



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 123041**

**TO: Georgia L Helmer  
Location: REM-2C15&2C18  
Art Unit: 1638  
Wednesday, June 02, 2004**

**Case Serial Number: 10/015637**

**From: Mary Jane Ruhl  
Location: Biotech-Chem Library  
Remsen 1-B55  
Phone: 571-272-2524**

**[maryjane.ruhl@uspto.gov](mailto:maryjane.ruhl@uspto.gov)**

### **Search Notes**

Examiner Helmer,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl  
Technical Information Specialist  
STIC  
CM-1, Rm. 6-A-06  
605-1155



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:15:29 ; Search time 4648 Seconds

(without alignments)  
10705.207 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148  
Sequence: 1 tagagatccttcacatagaanaa.....agagagtgatggtatgcga 1148Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

GenBank: 1: gb\_ba: 2: gb\_hg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pac: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_or: 22: em\_ov: 23: em\_pac: 24: em\_ph: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vl: 30: em\_hg\_hum: 31: em\_hg\_inv: 32: em\_hg\_other: 33: em\_hg\_mus: 34: em\_hg\_pln: 35: em\_hg\_rod: 36: em\_hg\_mam: 37: em\_hg\_vrt: 38: em\_sy: 39: em\_hgo\_hum: 40: em\_hgo\_mus: 41: em\_hgo\_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	1148	AX463281	AX463281 Sequence
2	1148	100.0	1832	AX463294	AX463294 Sequence
3	1144.8	99.7	1821	AX439309	AX439309 Sequence
4	1144.8	99.7	3900	PVARSX1	Z50202 P. vulgaris
5	1122	97.7	1122	AX463282	AX463282 Sequence
6	795.6	69.4	1866	AX463293	AX463293 Sequence
7	795	69.3	1872	AX463292	AX463292 Sequence
8	274.6	23.9	4564	PHVAC1A	ME8913 Phaseolus v
9	139.4	12.1	2288	AF193029	AF193029 Phaseolus
10	118.2	10.3	1768	PVPDLECI	X04609 P. vulgaris
11	109.2	9.5	22243	PPAVAR23A	I40609 Plasmodium
12	108.4	9.4	1395	PVPDLECI	X02408 Phaseolus v
13	108.4	9.4	1441	PHVDECA	K03288 P. vulgaris
14	95.8	8.3	246611	AC111404	AC111404 Rattus no
15	95.2	8.3	93791	AC138073	AC138073 Homo sapi
16	95	8.3	125933	AC146394	AC146394 Pan trogl
17	92	8.0	1689	PHYLECT	J01261 P. vulgaris
18	92	8.0	4846	AF325188	AF325188 Phaseolus
19	91.8	8.0	111861	AC069435	AC069435 Homo sapi
20	91.8	8.0	170627	PVPDLECI	AC125567 Rattus no
21	91.6	8.0	1392	PVPDLECI	AC117342 Rattus no
22	91.2	7.9	175544	AC117342	AC117342 Plasmodium
23	90.6	7.9	258658	AB014832	AB014832 Sequence
24	90	7.8	8056	PMAL8P1_12	AX599046 Sequence
25	90	7.8	110000	PMAL8P1_12	Continuation (13 o
26	89.8	7.8	810	AB001398	AB001398 Arabidops
27	89.8	7.8	14867	AL353783	AL353783 Human DNA
28	89.8	7.8	136688	AL353783	AL353783 Human DNA
29	89.6	7.8	103344	HS1100815	BS110640 Dario rer
30	89.6	7.8	141275	BX510640	BX510640 Zebrafish
31	89.6	7.8	171333	BX248097	AC127704 Rattus no
32	89.6	7.8	249867	AC127704	AL034560 Plasmodium
33	88.8	7.7	108908	PMAL8P8	BX088600 Dario rer
34	88.8	7.7	195620	BX088600	AB014820 Plasmodium
35	88.8	7.7	250029	AB014820	AB014820 Arabidops
36	88.6	7.7	250743	AB014836	U49822 Saccharomyc
37	87.4	7.6	1434	AB0192058	BX640469 Dario rer
38	87.4	7.6	2426	AB0192058	AB014830 Plasmodium
39	87.4	7.6	170880	AB014830	AB014830 Arabidops
40	87.4	7.6	250029	AB014830	AB014830 Arabidops
41	87.2	7.6	1453	AB0191978	AB0191978 Dario rer
42	87.2	7.6	164640	BX005461	BX005461 Dario rer
43	87.2	7.6	168558	BX322549	BX322549 Dario rer
44	87	7.6	3683	AX598999	AX598999 Sequence
45	87	7.6	136657	AC068600	AC068600 Homo sapi

## ALIGNMENTS

RESULT 1  
AX463281 1148 bp DNA linear PAT 15-JUL-2002  
LOCUS AX463281  
DEFINITION Sequence 1 from Patent WO0250295.  
ACCESSION AX463281  
VERSION AX463281.1 GI:21886232  
KEYWORDS  
SOURCE  
ORGANISM  
Phaseolus vulgaris  
Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE  
1 Oulmassov, T., Wang, Q., Dubois, P. and Liang, J.

TITLE Arcelin-5 promoter and uses thereof  
JOURNAL Patent: WO 0250295-A 1 27-JUN-2002;  
RENNESSEN LLC (US)  
FEATURES Location/Qualifiers  
source 1..1148  
/organism="Phaseolus vulgaris"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3885"

## ORIGIN

Query Match 100.0%; Score 1148; DB 6; Length 1148;  
Best local Similarity 100.0%; Pred. No. 8.4e-173;  
Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGATCTTCATAGAAAATGCTTATTTCCCATACACGACAAAGGGCAACGTTA 60  
DB 1 TAGGATCTTCATAGAAAATGCTTATTTCCCATACACGACAAAGGGCAACGTTA 60  
QY 61 ACAAACAATTAATGTTTCAATTTAGAGTAAGAGTAAGAGTAAGAGTAAGAGTA 120  
DB 61 ACAAACAATTAATGTTTCAATTTAGAGTAAGAGTAAGAGTAAGAGTAAGAGTA 120  
QY 121 AAAATGCTTATCTCTTTGTTTCTGTAATTAATTAATTAATTAATTAATTAAT 180  
DB 121 AAAATGCTTATCTCTTTGTTTCTGTAATTAATTAATTAATTAATTAATTAAT 180  
QY 181 AATTAATGTAATGTTTCTTCTAGTCAAGACACCTCAGACAGAGATTTCAGAAA 240  
DB 181 AATTAATGTAATGTTTCTTCTAGTCAAGACACCTCAGACAGAGATTTCAGAAA 240  
QY 241 CAATTTGTTAAACATCTTATTAAGAACTTTAGTTAGTTGAGTTAGATTAAACA 300  
DB 241 CAATTTGTTAAACATCTTATTAAGAACTTTAGTTAGTTGAGTTAGATTAAACA 300  
QY 301 AAAAAATTAACACGAGAAACAAATTAACCACTCCGTCAGGTTATCAATAGATGA 360  
DB 301 AAAAAATTAACACGAGAAACAAATTAACCACTCCGTCAGGTTATCAATAGATGA 360  
QY 361 AATGTTTGAATGATTAATTAATTAACACACACAAATAATACATTAATTAACAAATAT 420  
DB 361 AATGTTTGAATGATTAATTAATTAACACACACAAATAATACATTAATTAACAAATAT 420  
QY 421 GTTATACATATTTTGTAAATAAAGTTTGTAAAGTTTGTAAAGTTTGTAAAGTTTGT 480  
DB 421 GTTATACATATTTTGTAAATAAAGTTTGTAAAGTTTGTAAAGTTTGTAAAGTTTGT 480  
QY 481 AGTTATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540  
DB 481 AGTTATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540  
QY 541 CAGATCTCTCAACCTGTGTGATCATTTAGTCAATGATGATGATGATGATGATGATG 600  
DB 541 CAGATCTCTCAACCTGTGTGATCATTTAGTCAATGATGATGATGATGATGATGATG 600  
QY 601 ACAACAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 660  
DB 601 ACAACAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 660  
QY 661 AAAAAAGGAAATCAATTAATTAATTTTGAATTTCCCAATGACACAACTCACATGAC 720  
DB 661 AAAAAAGGAAATCAATTAATTAATTTTGAATTTCCCAATGACACAACTCACATGAC 720  
QY 721 GCTGCACTCAAGCTCCCTCTCTCCACACAGTCTCATGTCACCTTCTGCTTGGCTT 780  
DB 721 GCTGCACTCAAGCTCCCTCTCTCCACACAGTCTCATGTCACCTTCTGCTTGGCTT 780  
QY 781 TTCACTATGACACAACTGCGCATGATGTCAGAGCTCTCTCTCTCTCTCTCTCTCT 840  
DB 781 TTCACTATGACACAACTGCGCATGATGTCAGAGCTCTCTCTCTCTCTCTCTCTCT 840  
QY 841 GACACCACTGGGCGATGATGCTGCGCACTGAGCTCTCTCTCTCTCTCTCTCTCT 900  
DB 841 GACACCACTGGGCGATGATGCTGCGCACTGAGCTCTCTCTCTCTCTCTCTCTCT 900

QY 901 TGGCCATGACACAGCGACCTGACACATCTCTCACTTCCCATGCTGCAACCG 960  
DB 901 TGGCCATGACACAGCGACCTGACACATCTCTCACTTCCCATGCTGCAACCG 960  
QY 961 CTCTCTCCATTAATATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
DB 961 CTCTCTCCATTAATATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
QY 1021 CGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
DB 1021 CGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
QY 1081 AAGAAAAAGTTGAAAGATTTTGTGATTTGTGATTAATTAAGAAAGAGATGATG 1140  
DB 1081 AAGAAAAAGTTGAAAGATTTTGTGATTTGTGATTAATTAAGAAAGAGATGATG 1140  
QY 1141 TTAATGCA 1148  
DB 1141 TTAATGCA 1148

RESULT 2  
AX463294 1832 bp DNA linear PAT 15-JUL-2002  
LOCUS AX463294  
DEFINITION Sequence 14 from Patent WO0250295.  
ACCESSION AX463294  
VERSION AX463294.1 GI:21886245  
KEYWORDS  
SOURCE  
ORGANISM  
Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.

REFERENCE  
AUTHORS Ouljassov,T., Wang,Q., Dubois,P. and Liang,J.  
TITLE Arcelin-5 promoter and uses thereof  
JOURNAL Patent: WO 0250295-A 14 27-JUN-2002;  
RENNESSEN LLC (US)  
FEATURES Location/Qualifiers  
source 1..1832  
/organism="Phaseolus vulgaris"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3885"

## ORIGIN

Query Match 100.0%; Score 1148; DB 6; Length 1832;  
Best local Similarity 100.0%; Pred. No. 7.7e-173;  
Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGATCTTCATAGAAAATGCTTATTTCCCATACACGACAAAGGGCAACGTTA 60  
DB 659 TAGGATCTTCATAGAAAATGCTTATTTCCCATACACGACAAAGGGCAACGTTA 718  
QY 61 ACAAACAATTAATGTTTCAATTTAGAGTAAGAGTAAGAGTAAGAGTAAGAGTA 120  
DB 719 ACAAACAATTAATGTTTCAATTTAGAGTAAGAGTAAGAGTAAGAGTAAGAGTA 120  
QY 121 AAAATGCTTATCTCTTTGTTTCTGTAATTAATTAATTAATTAATTAATTAAT 180  
DB 779 AAAATGCTTATCTCTTTGTTTCTGTAATTAATTAATTAATTAATTAATTAAT 180  
QY 181 AATTAATGTAATGTTTCTTCTAGTCAAGACACCTCAGACAGAGATTTCAGAAA 240  
DB 839 AATTAATGTAATGTTTCTTCTAGTCAAGACACCTCAGACAGAGATTTCAGAAA 240  
QY 241 CAATTTGTTAAACATCTTATTAAGAACTTTAGTTAGTTGAGTTAGATTAAACA 300  
DB 899 CAATTTGTTAAACATCTTATTAAGAACTTTAGTTAGTTGAGTTAGATTAAACA 300  
QY 301 AAAAAATTAACACGAGAAACAAATTAACCACTCAGGTTATCATTAAGATGA 360

Db 959 AAAAAATTCACAGAGAAACAAATAACCCCTACCGTACGTTATCATAGATGA 1018  
 Qy 361 AATGTTTGAATATCATTAATAATTAACACACACAAAAATACATCTAATTAACAAATAT 420  
 Db 1019 AATGTTTGAATATCATTAATAATTAACACACAAAAATACATCTAATTAACAAATAT 1078  
 Qy 421 GTTATACATATATTTTGTAAAAACTTAAAGTTTTCAAAATCTAATTAACATATAG 480  
 Db 1079 GTTATACATATATTTTGTAAAAACTTAAAGTTTTCAAAATCTAATTAACATATAG 1138  
 Qy 481 AGTTATAGAAATACAAATATTTTAAAAATATATATTTTAAAAAATCTTCAAGTCAT 540  
 Db 1139 AGTTATAGAAATACAAATATTTTAAAAATATATATTTTAAAAAATCTTCAAGTCAT 1198  
 Qy 541 CAGATCTCTCACACCTGTGTGATCTTTTATCATGTATATGTATACATCTTGTATTC 600  
 Db 1199 CAGATCTCTCACACCTGTGTGATCTTTTATCATGTATATGTATACATCTTGTATTC 1258  
 Qy 601 ACAACAGAGTAAAT 660  
 Db 1259 ACAACAGAGTAAAT 1318  
 Qy 661 AAAAAAGGAAAAATCAATATAGATTTTGTATTTTCCCAATGACACAACTCACATGAC 720  
 Db 1319 AAAAAAGGAAAAATCAATATAGATTTTGTATTTTCCCAATGACACAACTCACATGAC 1378  
 Qy 721 GCTGCACCTGACCTGCT 780  
 Db 1379 GCTGCACCTGACCTGCT 1438  
 Qy 781 TTCATATGACACAACTGCCATGATGTCACGTCAGTCTCTCTCTCTCTCTCTCTCT 840  
 Db 1439 TTCATATGACACAACTGCCATGATGTCACGTCAGTCTCTCTCTCTCTCTCTCTCT 1498  
 Qy 841 GACACCACTGGGATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCTCTCT 900  
 Db 1499 GACACCACTGGGATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCTCTCT 1558  
 Qy 901 TGCCCATGACACCTGCCATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCT 960  
 Db 1559 TGCCCATGACACCTGCCATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCT 1618  
 Qy 961 CTTCTCTCAATAAAT 1020  
 Db 1619 CTTCTCTCAATAAAT 1078  
 Qy 1021 CGTGCATGACCTGCCATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCTCT 1080  
 Db 1679 CGTGCATGACCTGCCATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCTCT 1738  
 Qy 1081 AAAAAAGGAAAAATCAATATAGATTTTGTATTTTCCCAATGACACAACTCACATGAC 1140  
 Db 1739 AAAAAAGGAAAAATCAATATAGATTTTGTATTTTCCCAATGACACAACTCACATGAC 1798  
 Qy 1141 TTATATGCA 1148  
 Db 1799 TTATATGCA 1806

RESULT 3  
 AX343909 1821 bp DNA linear PAT 01-FEB-2002  
 LOCUS AX343909 Sequence 1 from Patent WO0200899.  
 ACCESSION AX343909  
 VERSION AX343909.1 GI:18491955  
 KEYWORDS  
 ORGANISM Phaseolus vulgaris  
 Phaseolus vulgaris  
 Bkaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Phaseolus.  
 REFERENCE 1

AUTHORS Angenon, G., de Jaeger, G., Goossens, A. and Depicker, A.  
 TITLE Heterologous gene expression in plants  
 JOURNAL Patent: WO 0200899-A 1 03-JAN-2002;  
 Vitals Interuniversitair Instituut voor Biotechnologie vzw. (BE)  
 FEATURES  
 source 1. 1821  
 /organism="Phaseolus vulgaris"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3885"

Query Match 99.74; Score 1144.8; DB 6; Length 1821;  
 Best Local Similarity 99.83; Pred. No. 2.5e-172;  
 Matches 1146; Conservatives 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGATCCTTCAATAGAAATATGTTTATTTCTCATGACACGACAAAGGGCAACAGTTA 60  
 Db 674 TGGATCCTTCAATAGAAATATGTTTATTTCTCATGACACGACAAAGGGCAACAGTTA 733  
 Qy 61 ACAAACAAATTTATGTTTCATTTGAGATTAGGAAGGTAGGAAGAAAGATTAAAA 120  
 Db 734 ACAAACAAATTTATGTTTCATTTGAGATTAGGAAGGTAGGAAGAAAGATTAAAA 793  
 Qy 121 AATATGCTTATCTCTTTGTTTCTGTAATATATATATATATATATATATATATATAT 180  
 Db 794 AATATGCTTATCTCTTTGTTTCTGTAATATATATATATATATATATATATATATAT 853  
 Qy 181 AATATGTTAAT 240  
 Db 854 AATATGTTAAT 913  
 Qy 241 CAATTTGTTAAACATCTTATATAGAACTTTTATAGTCTTGAAGTTAGATTAACA 300  
 Db 914 CAATTTGTTAAACATCTTATATAGAACTTTTATAGTCTTGAAGTTAGATTAACA 973  
 Qy 301 AAAAAATTTACACAGAGAAACAAATTAACCACTACCGTACAGTTATCATAGATGA 360  
 Db 974 AAAAAATTTACACAGAGAAACAAATTAACCACTACCGTACAGTTATCATAGATGA 1033  
 Qy 361 AATGTTTGAATATCAATTAATATTAACACACAAATATACATTAATTAACAATATAT 420  
 Db 1034 AATGTTTGAATATCAATTAATATTAACACACAAATATACATTAATTAACAATATAT 1093  
 Qy 421 GTTATACATATATTTTGTAAAAACTTGAAGTTTCAAAATCTAATATATATATATAT 480  
 Db 1094 GTTATACATATATTTTGTAAAAACTTGAAGTTTCAAAATCTAATATATATATATAT 1153  
 Qy 481 AGTTATAGAAATACAAATATTTAAAAATATATATATATATATATATATATATATAT 540  
 Db 1154 AGTTATAGAAATACAAATATTTAAAAATATATATATATATATATATATATATATAT 1213  
 Qy 541 CAGATCCTCTCACACCTGTGTGATCTTTATGTCAGTATATGTATCAATCTAGATTC 600  
 Db 1214 CAGATCCTCTCACACCTGTGTGATCTTTATGTCAGTATATGTATCAATCTAGATTC 1273  
 Qy 601 ACAACAGAGTAAAT 660  
 Db 1274 ACAACAGAGTAAAT 1333  
 Qy 661 AAAAAAGGAAAAATCAATATAGATTTTGTATTTTCCCAATGACACAACTCACATGAC 720  
 Db 1334 AAAAAAGGAAAAATCAATATAGATTTTGTATTTTCCCAATGACACAACTCACATGAC 1393  
 Qy 721 GCTGCACCTGACCTGCT 780  
 Db 1394 GCTGCACCTGACCTGCT 1453  
 Qy 781 TTCATATGACACAACTGCCATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCT 840  
 Db 1454 TTCATATGACACAACTGCCATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCT 1513  
 Qy 841 GACACCACTGGGATGATGTCAGTCTGCCATGATGTCAGTCTCTCTCTCTCTCTCTCTCT 900

[illegible]

Db 854 AATAATGTAATAGGTTTCTAGTCATAGACCACTAGAGACAGATTTCAGAAAA 913  
 Qy 241 CAATTTGTTAAACATCTATTAGAAAATTTAGTTAACTCTGGAATGATTAACA 300  
 Db 914 CAATTTGTTAAACATCTATTAGAAAATTTAGTTAACTCTGGAATGATTAACA 973  
 Qy 301 AAAAAATTATGACACGAGAAACCATTAACCCACTAGCTGAGTTATCATPAGATCA 360  
 Db 974 AAAAAAGTACACACGAGAAACCATTAACCCACTAGCTGAGTTATCATPAGATCA 1033  
 Qy 361 AATGTTGATATCATTAATTAATTAACACAAAAATACATCTAATTAACAATAT 420  
 Db 1034 AATGTTGATATCATTAATTAATTAACACAAAAATACATCTAATTAACAATAT 1093  
 Qy 421 GTTATACATATATTTTGTAAAAAATTAGAGTTTCAAAAACATCTAATTAACAT 480  
 Db 1094 GTTATACATATATTTTGTAAAAAATTAGAGTTTCAAAAACATCTAATTAACAT 1153  
 Qy 481 AGTTTATAGAAATACAAATTTTAAAAATATATTTTAAAAAATCTTAAGCAT 540  
 Db 1154 AGTTTATAGAAATACAAATTTTAAAAATATATTTTAAAAAATCTTAAGCAT 1213  
 Qy 541 CAGATCTCTCACACCTGTGTGATCATTTAGTCATGTAGTACATCAATCTTGTAG 600  
 Db 1214 CAGATCTCTCACACCTGTGTGATCATTTAGTCATGTAGTACATCAATCTTGTAG 1273  
 Qy 601 ACAACAGAGTAAATTAATTAAGATTAACCTAGGAAATATATATTAATTAACA 660  
 Db 1274 ACAACAGAGTAAATTAATTAAGATTAACCTAGGAAATATATATTAATTAACA 1333  
 Qy 661 AAAAAAGGAAATATCAATTTAGATTTTCCCATGACACCACTGACATGAC 720  
 Db 1334 AAAAAAGGAAATATCAATTTAGATTTTCCCATGACACCACTGACATGAC 1393  
 Qy 721 GGTGCACTCTCAGCTCTCTCTCTCCACATGTCTCATGTCTTCACTTGTGCTTT 780  
 Db 1394 GGTGCACTCTCAGCTCTCTCTCTCTCCACATGTCTCATGTCTTCACTTGTGCTTT 1453  
 Qy 781 TTCACTATGACACCACTGCGCATGATGTGCGACGAGCTCTCTCTCTCTCTGAT 840  
 Db 1454 TTCACTATGACACCACTGCGCATGATGTGCGACGAGCTCTCTCTCTCTGAT 1513  
 Qy 841 GACACCACTGCGCATGATGTGCGACGAGCTCTCTCTCTCTCTGATGAGCTTAC 900  
 Db 1514 GACACCACTGCGCATGATGTGCGACGAGCTCTCTCTCTCTCTGATGAGCTTAC 1573  
 Qy 901 TGCCCATGACACCTGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
 Db 1574 TGCCCATGACACCTGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1633  
 Qy 961 CT 1020  
 Db 1634 CT 1693  
 Qy 1021 CGTGAATGACATTCGATCTGTTTAAATTTGTAATTTGAGTTGAATTAATTA 1080  
 Db 1694 CGTGAATGACATTCGATCTGTTTAAATTTGTAATTTGAGTTGAATTAATTA 1753  
 Qy 1081 AAAAAAAGTTGGAAGAATTTTGATTTGTTGATTAATTAAGAGAAGAGTATG 1140  
 Db 1754 AAAAAAAGTTGGAAGAATTTTGATTTGTTGATTAATTAAGAGAAGAGTATG 1813  
 Qy 1141 TTAATGCA 1148  
 Db 1814 TTAATGCA 1821

RESULT 5  
 AX463282  
 LOCUS  
 DEFINITION Sequence 2 from Patent WO0250295.  
 AX463282  
 ACCESSION  
 1122 bp DNA linear PAT 15-JUL-2002

VERSION AX463282.1 GI:2186233  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Phaseolus vulgaris  
 Phaseolus vulgaris  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Phaseolus.  
 REFERENCE  
 1 Oulmassov, T., Wang, O., Dubois, P. and Liang, J.  
 Arcelin-5 promoter and uses thereof  
 Patent: WO 0250295-A 27-JUN-2002;  
 REMSENSEN LLC (US)  
 FEATURES  
 source location/Qualifiers  
 1..1122  
 /organism="Phaseolus vulgaris"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3855"  
 ORIGIN  
 Query Match 97.7%; Score 1122; DB 6; Length 1122;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-168; Mismatches 0; Gaps 0;  
 Matches 1122; Conservative 0; Nismatches 0; Indels 0; Gaps 0;  
 Qy 3 GGATCCTTCATAGAAAATGTGTATTTCTCATCAACACAAAGGGGCAACGTTAAC 62  
 Db 1 GGATCCTTCATAGAAAATGTGTATTTCTCATCAACACAAAGGGGCAACGTTAAC 60  
 Qy 63 AAAAAAATTAATGTTATGATTTAGATTAAGAGAGTAAGAGAAAAAGTTAAAAA 122  
 Db 61 AAAAAAATTAATGTTATGATTTAGATTAAGAGAGTAAGAGAAAAAGTTAAAAA 120  
 Qy 123 AATGCTTATCTCTTGTCTGTATTAATTAATTAAGAGAGTTAACTTTAATATA 182  
 Db 121 AATGCTTATCTCTTGTCTGTATTAATTAATTAAGAGAGTTAACTTTAATATA 180  
 Qy 183 TAATTTATTAATGTTTCTGTATTAATTAATTAAGAGAGTTAACTTTAATATA 242  
 Db 181 TAATTTATTAATGTTTCTGTATTAATTAATTAAGAGAGTTAACTTTAATATA 240  
 Qy 243 AATTTGTTAAATCTTTATTAAGAACTTTTATAGTCTTGAATTAATTAACAA 302  
 Db 241 AATTTGTTAAATCTTTATTAAGAACTTTTATAGTCTTGAATTAATTAACAA 300  
 Qy 303 AAAAAATTAACACGAGAAACCAATTAACCCACTACCGTACGTTATCATTAAGATGAA 362  
 Db 301 AAAAAATTAACACGAGAAACCAATTAACCCACTACCGTACGTTATCATTAAGATGAA 360  
 Qy 363 TGTTTTGAATATTAATTAATTAACACCAAAATTAACCTAATTAATTAACATATATGT 422  
 Db 361 TGTTTTGAATATTAATTAATTAACACCAAAATTAACCTAATTAATTAACATATATGT 420  
 Qy 423 TATACATATATTTTGTAAAAAATTAGAGTTTTCAAAAATTCCTAATTAATTAAG 482  
 Db 421 TATACATATATTTTGTAAAAAATTAGAGTTTTCAAAAATTCCTAATTAATTAAG 480  
 Qy 483 TTTATGAATATCAATATTTTAAAAATTAATTTTAAAAAATCTTAAAGTCATCA 542  
 Db 481 TTTATGAATATCAATATTTTAAAAATTAATTTTAAAAAATCTTAAAGTCATCA 540  
 Qy 543 GATCCTCTCACACCTGTGTGATCATTTAGTCATGTAGTACATCAATCTTGTAGTAC 602  
 Db 541 GATCCTCTCACACCTGTGTGATCATTTAGTCATGTAGTACATCAATCTTGTAGTAC 600  
 Qy 603 AACAGATTAATTAATTAAGATTAACCTAGGAAATATATTAATTAATTAATTAATTA 662  
 Db 601 AACAGATTAATTAATTAAGATTAACCTAGGAAATATATTAATTAATTAATTAATTA 660  
 Qy 663 AAAAAAATCAATTAATTAATTTTGTATTTTCCCATGACACCACTACCATGACGC 722  
 Db 661 AAAAAAATCAATTAATTAATTTTGTATTTTCCCATGACACCACTACCATGACGC 720  
 Qy 723 TGCCACCTACGCTCCCTCTCTCCACACATGTCTCATGTCTTGTGATTTGCTTTT 782



Accession	Version	AX463292
KeyWords	AX463292.1 GI:21886243	
Source	Phaeoelus vulgaris	
Organism	Phaeoelus vulgaris	
Reference	Phaeoelus vulgaris	
Authors	Phaeoelus vulgaris	
Title	Phaeoelus vulgaris	
Journal	Phaeoelus vulgaris	
Features	Phaeoelus vulgaris	
Source	Phaeoelus vulgaris	
Origin	Phaeoelus vulgaris	
Query Match	69.3%; Score 795; DB 6; Length 1872;	
Best Local Similarity	85.7%; Pred. No. 7.8e-117;	
Matches 1043; Conservative	0; Mismatches 90; Indels 84; Gaps 11;	
1	Phaeoelus vulgaris	
2	Phaeoelus vulgaris	
3	Phaeoelus vulgaris	
4	Phaeoelus vulgaris	
5	Phaeoelus vulgaris	
6	Phaeoelus vulgaris	
7	Phaeoelus vulgaris	
8	Phaeoelus vulgaris	
9	Phaeoelus vulgaris	
10	Phaeoelus vulgaris	
11	Phaeoelus vulgaris	
12	Phaeoelus vulgaris	
13	Phaeoelus vulgaris	
14	Phaeoelus vulgaris	
15	Phaeoelus vulgaris	
16	Phaeoelus vulgaris	
17	Phaeoelus vulgaris	
18	Phaeoelus vulgaris	
19	Phaeoelus vulgaris	
20	Phaeoelus vulgaris	
21	Phaeoelus vulgaris	
22	Phaeoelus vulgaris	
23	Phaeoelus vulgaris	
24	Phaeoelus vulgaris	
25	Phaeoelus vulgaris	
26	Phaeoelus vulgaris	
27	Phaeoelus vulgaris	
28	Phaeoelus vulgaris	
29	Phaeoelus vulgaris	
30	Phaeoelus vulgaris	
31	Phaeoelus vulgaris	
32	Phaeoelus vulgaris	
33	Phaeoelus vulgaris	
34	Phaeoelus vulgaris	
35	Phaeoelus vulgaris	
36	Phaeoelus vulgaris	
37	Phaeoelus vulgaris	
38	Phaeoelus vulgaris	
39	Phaeoelus vulgaris	
40	Phaeoelus vulgaris	
41	Phaeoelus vulgaris	
42	Phaeoelus vulgaris	
43	Phaeoelus vulgaris	
44	Phaeoelus vulgaris	
45	Phaeoelus vulgaris	
46	Phaeoelus vulgaris	
47	Phaeoelus vulgaris	
48	Phaeoelus vulgaris	
49	Phaeoelus vulgaris	
50	Phaeoelus vulgaris	
51	Phaeoelus vulgaris	
52	Phaeoelus vulgaris	
53	Phaeoelus vulgaris	
54	Phaeoelus vulgaris	
55	Phaeoelus vulgaris	
56	Phaeoelus vulgaris	
57	Phaeoelus vulgaris	
58	Phaeoelus vulgaris	
59	Phaeoelus vulgaris	
60	Phaeoelus vulgaris	
61	Phaeoelus vulgaris	
62	Phaeoelus vulgaris	
63	Phaeoelus vulgaris	
64	Phaeoelus vulgaris	
65	Phaeoelus vulgaris	
66	Phaeoelus vulgaris	
67	Phaeoelus vulgaris	
68	Phaeoelus vulgaris	
69	Phaeoelus vulgaris	
70	Phaeoelus vulgaris	
71	Phaeoelus vulgaris	
72	Phaeoelus vulgaris	
73	Phaeoelus vulgaris	
74	Phaeoelus vulgaris	
75	Phaeoelus vulgaris	
76	Phaeoelus vulgaris	
77	Phaeoelus vulgaris	
78	Phaeoelus vulgaris	
79	Phaeoelus vulgaris	
80	Phaeoelus vulgaris	
81	Phaeoelus vulgaris	
82	Phaeoelus vulgaris	
83	Phaeoelus vulgaris	
84	Phaeoelus vulgaris	
85	Phaeoelus vulgaris	
86	Phaeoelus vulgaris	
87	Phaeoelus vulgaris	
88	Phaeoelus vulgaris	
89	Phaeoelus vulgaris	
90	Phaeoelus vulgaris	
91	Phaeoelus vulgaris	
92	Phaeoelus vulgaris	
93	Phaeoelus vulgaris	
94	Phaeoelus vulgaris	
95	Phaeoelus vulgaris	
96	Phaeoelus vulgaris	
97	Phaeoelus vulgaris	
98	Phaeoelus vulgaris	
99	Phaeoelus vulgaris	
100	Phaeoelus vulgaris	
101	Phaeoelus vulgaris	
102	Phaeoelus vulgaris	
103	Phaeoelus vulgaris	
104	Phaeoelus vulgaris	
105	Phaeoelus vulgaris	
106	Phaeoelus vulgaris	
107	Phaeoelus vulgaris	
108	Phaeoelus vulgaris	
109	Phaeoelus vulgaris	
110	Phaeoelus vulgaris	
111	Phaeoelus vulgaris	
112	Phaeoelus vulgaris	
113	Phaeoelus vulgaris	
114	Phaeoelus vulgaris	
115	Phaeoelus vulgaris	
116	Phaeoelus vulgaris	
117	Phaeoelus vulgaris	
118	Phaeoelus vulgaris	
119	Phaeoelus vulgaris	
120	Phaeoelus vulgaris	
121	Phaeoelus vulgaris	
122	Phaeoelus vulgaris	
123	Phaeoelus vulgaris	
124	Phaeoelus vulgaris	
125	Phaeoelus vulgaris	
126	Phaeoelus vulgaris	
127	Phaeoelus vulgaris	
128	Phaeoelus vulgaris	
129	Phaeoelus vulgaris	
130	Phaeoelus vulgaris	
131	Phaeoelus vulgaris	
132	Phaeoelus vulgaris	
133	Phaeoelus vulgaris	
134	Phaeoelus vulgaris	
135	Phaeoelus vulgaris	
136	Phaeoelus vulgaris	
137	Phaeoelus vulgaris	
138	Phaeoelus vulgaris	
139	Phaeoelus vulgaris	
140	Phaeoelus vulgaris	
141	Phaeoelus vulgaris	
142	Phaeoelus vulgaris	
143	Phaeoelus vulgaris	
144	Phaeoelus vulgaris	
145	Phaeoelus vulgaris	
146	Phaeoelus vulgaris	
147	Phaeoelus vulgaris	
148	Phaeoelus vulgaris	
149	Phaeoelus vulgaris	
150	Phaeoelus vulgaris	
151	Phaeoelus vulgaris	
152	Phaeoelus	

QY	668	GGAAAAATCAATATGAAATTTTATTTATTTATTTCCCAATGACAAATCAGCATCAGCTGCCA	727
Db	1350	GGAAAAATCAATATGAAATTTTATTTATTTATTTCCCAATGACAAATCAGCATCAGCTGCCA	1409
QY	728	CTTCAAGCTCCCTCTCTCCACAATGTCATATGTCATTTCAGCTTTGGCTTTTTCATCTA	787
Db	1410	CTTCAAGCTCCCTCTCTCCACAATGTCATATGTCATTTCAGCTTTGGCTTTTTCATCTA	1469
QY	788	TGACACAACTCGGCATGTCAGATGTTGGACACGAGCTGCTCTCTCTGCCATGATGACACCA	847
Db	1470	GGAGACAACTCGGCATGTCAGATGTTGGACACGAGCTGCTCTCTCTGCCATGATGACACCA	1529
QY	848	CTGGGCATGATGCTGTCGCCACTCAGCTCCACCTCTCTCTCATATGAGCTTACTGGCCAT	907
Db	1530	CTGGGCATGATGATGTCACCTCAGCTCCACCTCTCTCATATGAGCTTACTGGCCAT	1589
QY	908	GCAACACGTCACCTCAGACATCTCTCTCACTTCCTCATGTCATGCAACGACCAACCGCTTCTCT	967
Db	1590	GCAACACGTCACCTCAGACATCTCTCTCACTTCCTCATGTCATGCAACGACCAACCGCTTCTCT	1649
QY	968	CCATTAATATCTATTTTAATTTAACTAATATTTTCAATATACCTTTTTCATGACGTCGAT	1027
Db	1650	CTATTAATATCTCTTTAAATTTAACTAATATTTTCAATATACCTTTTTCATGACGTCGAT	1709
QY	1028	GCATTGCGCATGCTGTTGTTTAATTAATGTTTAATTGGAGTT-----GAAATAT	1073
Db	1710	GCATTGCGCATGCTGTTGTTTAATTAATGTTTAATTGTAATTAATTTTCTCCTCAAAATAT	1769
QY	1074	AAATGGAAGAAAAAAGTTGGA--AAATTTTGCATTTGTTGTTAATTAATGAGAGA	1131
Db	1770	ATTATTAAGATTAATGATTCGATTTGATTTTGGATTTGTTGTTAATTAATGAGAGA	1829
QY	1132	GAGTGATGCTTAATGCA	1148
Db	1830	GAGTGATGCTTAATGCA	1846
RESULT 8			
PHVARCIA		4564 bp	DNA linear PLN 27-APR-1993
LOCUS			
DEFINITION	Phaseolus vulgaris arcelin (arc) gene, complete cds.		
ACCESSION	M68913		
VERSION	M68913.1 GI:169313		
KEYWORDS	arcelin; lectin-like seed protein.		
SOURCE	Phaseolus vulgaris		
ORGANISM	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.		
REFERENCE	1 (bases 1 to 4564)		
AUTHORS	Anthony, D.L., Vonder Haar, R.A. and Hall, T.C.		
TITLE	Nucleotide sequence of a genomic clone encoding arcelin, a lectin-like seed protein from Phaseolus		
JOURNAL	Plant Physiol. 97, 839-840 (1991)		
COMMENT	Original source text: Phaseolus vulgaris (library: Lambda Zap II) young plant leaf DNA.		
FEATURES			
Source	Location/Qualifiers		
	1..4564		
	/organism="Phaseolus vulgaris"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:3885"		
	/tissue_type="leaf"		
	/dev_stage="young plant"		
	/tissue_1ib="Lambda Zap II"		
	3332..4172		
	/gene="arc"		
	3332..3338		
	/gene="arc"		
	3361..3364		
	/gene="arc"		
	/note="cap region"		
	3375..4172		
CDS			



```

/gene="arc"
/codon_start=1
/product="arcelin"
/protein_id="AAA33753.1"
/db_xref="GI:169314"
/translation="MASSNLLTALFLVLLTHANSNDASFWETENKNTLLQGDAT
VSSGHLTLNVKNEDESMGRFAPQIDRTIDNLSPTNFTIRINKNTENS
AVGLAFLVPGVSRPLKGRVYGLFNTTIVSDADAVAVPTVSRLIDVNSLRPI
ATESCNFNGNGEKAEVRIYDPSKNDLVSLLYPSSEKCHVASFVPLEKEVDWVS
VGSATSGSKKETETETHNVLSFSSNFIFNGKXKSERNNILLNKL"
3375..3437
/gene="arc"
/motif="4169
/gene="arc"
/product="arcelin"

```

## ORIGIN

```

Query Match      23.9%; Score 274.6; DB 8; Length 4564;
Best Local Similarity 91.6%; Pred. No. 2.5e-34;
Matches 316; Conservative 0; Mismatches 19; Indels 10; Gaps 2;

QY 805 CATGTCACACGACGACGCTCTCTCTCCATGATGACACACCTGCGATGCTGC 864
DB 3028 CACTGACGACGACGCTCTCTCTCCATGATGACACACCTGCGATGCTGC 3087
QY 865 CACTGACGCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 924
DB 3088 CACTGACGCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3138
QY 925 CA-CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 983
DB 3139 CACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3198
QY 984 AAATTAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1043
DB 3199 AAATTAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3258
QY 1044 TTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1103
DB 3259 TTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3318
QY 1104 GCATTGCTGCTGTAATAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1148
DB 3319 GCATTGCTGCTGTAATAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3363

```

```

RESULT 9
AF193029 2288 bp DNA linear PLN 10-JAN-2000
LOCUS Phaseolus vulgaris arcelin 5c gene, complete cds.
DEFINITION AF193029
ACCESSION AF193029
VERSION AF193029.1 GI:6684755
KEYWORDS
SOURCE Phaseolus vulgaris
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1 (bases 1 to 2288)
AUTHORS Gerhardt,I.R. and Grossi de Sa,M.F.
TITLE Molecular characterization of a new arcelin-5 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2288)
AUTHORS Gerhardt,I.R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Embriapa/Cenargen, SAIN-Parque Rural-Final
W5 Norte, Brasilia, DF 70770-900, Brazil
FEATURES
source
1..2288
/organism="Phaseolus vulgaris"
/mol_type="genomic DNA"
/cultivar="G02771"

```

## ORIGIN

```

Query Match      12.1%; Score 139.4; DB 8; Length 2288;
Best Local Similarity 85.8%; Pred. No. 8e-13;
Matches 169; Conservative 0; Mismatches 21; Indels 7; Gaps 1;

QY 959 CGCTTCTCCATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1018
DB 393 CACTATCTCCATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 452
QY 1019 GACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
DB 453 GACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 512
QY 1072 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1131
DB 513 TCAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 572
QY 1132 GAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148
DB 573 GAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 589

```

```

RESULT 10
PVPDLEC1 1768 bp DNA linear PLN 12-SEP-1993
LOCUS P. vulgaris cv. Pinto pseudogene plect for phytohemagglutinin
DEFINITION PVPDLEC1
ACCESSION X04660
VERSION X04660.1 GI:21020
KEYWORDS plect gene; phytohemagglutinin; pseudogene.
SOURCE Phaseolus vulgaris
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Voelker,T.A., Staewick,P. and Chrispeels,M.J.
TITLE Molecular analysis of two phytohemagglutinin genes and their
expression in Phaseolus vulgaris cv. Pinto, a lectin-deficient
cultivar of the bean
JOURNAL EMO J. 5, 3075-3082 (1986)
COMMENT See also <X04659> For plect2 gene.
Several inverted repeats are described in the 5' upstream region of
the PNA gene.
None of the 3 possible reading frames allow the translation of its
message into a complete PNA polypeptide. The PNA translational
start codon (pos. 708-710) starts a RF which codes for a truncated
50AA polypeptide with a N-terminal sequence completely identical to
the dlect gene product, but is out of register after 10 codons due
to a single bp deletion (pos. 739/740).
location/Qualifiers
1..1768
/organism="Phaseolus vulgaris"
/mol_type="genomic DNA"
/strain="Pinto U111"

```

variation	misc_feature	ORIGIN
		<p>promoter</p> <p>CDS</p> <p>variation</p> <p>misc_feature</p> <p>ORIGIN</p>
		<p>Query Match</p> <p>Best Local Similarity</p> <p>Matches 352; Conservative</p>
		<p>10.3%; Score 118.2; DB 8; Length 1768;</p> <p>64.4%; Pred. No. 1.9e-09;</p> <p>0; Mismatches 125; Indels 70; Gaps 9;</p>
		<p>636 ATATATATTAATATATACATTTAAATTAAGAAATCAAAATAGAAATTTTGATTC 695</p> <p>186 ATTTATTTATTTAAATTAATTTATTTTAACTAACAAGGTTTGTGACTGTGGCGGTT 245</p> <p>696 CCACATGACACAATCACAATGACAGCGTGCACCTGACCTCCCTCTCCACATATGC 755</p> <p>246 TACAGTGCAGCAATGCAAGATGAGTGCACCTGACCTGCTCTTTCACACGTATC 305</p> <p>756 -TCATGTCACTTTCGACTTTGGCTTTTCACTATGACACAATCTGCCATGCAT---GTT 810</p> <p>306 TTCCTCTCACTTCCACATGACAGCTTTCACACTATGACATCACTGGTCAATGACAGCT 365</p> <p>811 GCCACGAGAGTCCGCTCTTCCCATGATGACACCACTGGGACATGATGCTGCCACTC 870</p> <p>366 GCCACCTGAGCTTCTCTCTCTTCCATCATGACAGACTGGCCATGATGATGCA---- 421</p> <p>871 AGCTCCCACTCTTTCATATATGAGCTTACGCTACGCACTGCCACTGACACTCC 930</p> <p>422 TGCTGCCACCTCAGCTC-----CCCTCCTCTTCAAC 450</p> <p>931 TCTCACTTCCCATTTGCTTACCTGCCAAACCGTCTCTCCATAATATCTATTTAAATTA 990</p> <p>451 CCGTGTCTCCAAAGCTACGTCGCCAACCGTCTCTTATATACATATCTTTAAATTA 509</p> <p>991 AA-CTAATATTTTCAAT-----ACTTTTGTATGACGTGATGATGCCATCGTTG 1042</p> <p>510 AACCTAATATTTCAATTTTAAAGTTTGTATGAGCGATGATGCTCCATCGTTG 569</p> <p>1043 TTATATATTTGTATTTTGAGTT-----GATATATTAATGAAGAAAGATTG- 1093</p> <p>570 CTATATCTATTTCTATATTTCTATTAATCTCAATATATATATATTAAGAAAGATAT 629</p> <p>1094 -----GAAAGATTTTGCATTGTGTGTGTATTAATAGAGAGAGTGTAGT 1141</p> <p>630 GAAGTAGAGAGAGAGAGAGTGTCTTGTGTGTATTAATAGAGAGAGTGTAGT 689</p> <p>1142 TAATGCA 1148</p> <p>690 TAATGCA 696</p>
		<p>RESULT 11</p> <p>PEAVAR23A</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p>
		<p>PEAVAR23A</p> <p>Plasmodium falciparum (strain FCRI3) variant-specific surface protein (var-2, var-3) genes, complete cds's.</p> <p>LA0609.1 GI:886376</p> <p>variant-specific surface protein.</p> <p>Plasmodium falciparum (malaria parasite P. falciparum)</p>

[illegible]

exon  
 /gene="var-3"  
 /number=1  
 8218..9092  
 /gene="var-3"  
 9093..>10322  
 /gene="var-3"  
 /number=2  
 11749..20741  
 /gene="var-2"  
 /gene="var-2"  
 /note="specific to blood stage; putative"  
 /codon\_start=1  
 /product="variant-specific surface protein"  
 /protein\_id="AA15398.1"  
 /db\_xref="GI:886378"  
 /translation="MVLQRAAGGAGGAGDIDKESAKHLIDSIGKYVDKRGDALQPSNG  
 KLGDTSLAIFKAEKQGTSEDPDCLNHEHTVTSIGYDNECDORPEKRSYFEG  
 AEDKSKTSGSNKQCAAPRILHCDQHLHIDKIDLTRENLADPEAKFPAB  
 SLKRYQYOLANSDVNINICTELASPRADIGDIVRGDILRGNDKREMLKRI  
 FKILYDNLDAHVOEHKDDKGTGKATYKLRNMBNRPYKATICTGAAGTFRQ  
 TCGITWTNEKCRPINDVPYFDYVLYRFBEMWDECRKQKATIDVKTGNY  
 SMLYSGKGLDCOETIRVLGHVIGSECSKSCWCRKWKIMDQCEELKQKKE  
 NEMLSKSKKQSTKXNYEGYDKFYLKLSBNVGLDKTELINERCOEBSNDG  
 KIDKRTVGGVGAIGGASDSNNNSNTPFSHQYCEPCGCVGLINEMKCKQK  
 KGRVNIPTGKHNVPVLSFGDEHKETIEKLEPCASNSSSKLTGEMCKYQK  
 EYECTLENKSKSEDPBEIOKTFNPFYMIHILNDSIEMDKINCKIEKQK  
 KNECTDGGCFQWIGKKEBMBIKGHFTQDGFSGNNYDPPVLYNLTIDELRD  
 ITRVAGNSQIKIGIKTLAKKTYADADATBQKATIDLLEYBSBBAKCKIQEIQ  
 PKKFTKVPQCYNNITYDALAGVAKQQLQEAQLDNRSSALKRVAQKSYNG  
 DPDPKKNLGGITQKSNALIGDSKPCNNKKEFPNVEGRNKGVEYMSHTLTPP  
 RROHCTSNLEHTKSTGLTSDKALHSLADVLLAAKKEGEDIKTLTENDNSIC  
 RTKYSRADIGDILRGTDINDGATGVONKLDIFSKITBELKQHPRENDNXY  
 TNSKRTKLSBDMDEANRDVWKAMCTPTKNGTQCATPHDVIPEGLNMYEAM  
 FQESQSLIEBLDCCSTTGKCNKCKAKCDKCCQCYETKIOPADONNELSNKYQ  
 ILVQAKIALINGSTKSTTKKDXKNVDFLOQLMBANVTGTPPEAPRPRPR  
 AATKSDVYETTAGYIHQEARTRCLQNFECNNNGNEBFAFSLTPEYGAACKNR  
 KASPEELGRSDSDHQTPPEDEVHSEGEDESDBEKEEVEVADGEKA  
 GAVOPEASPTTKDVWVPCDLYKELFSNVTLQKASTKYGKQKAMGNCIPIKTSN  
 DVNEDOGSSRRVYRSTPESGNSDKNGATCIPRRRLTYGKLEOMANKNETSVS  
 QGEATARGSAAPAGKESSGGKETPSDILTAFTESAVERFEFLMDRYKEMLLQ  
 KKAELQNGEDLISGDDPDPNPKLNGITPPRIMPTTGADYRDLIHGANTS  
 DSGTNSNNNNIVLEASGNKEDQKLEQILPKNGGTPLYPKSAQTPDRNMB  
 HAESIWKGMICALTYTEKNDPTSGADENKLEKDEYEFKFSFAHKGASTPST  
 YKTOYDEKVLIEDTSGAKTPSASDTPILSDPVLRPYEFELFGGONFCKEKL  
 KOKEBMDGSKKYSGGGDCDRDTSNEVASLDGSCNSCRFYKMLRKRKEY  
 DKQNAAYKOKTKYRBSGKAGLNDHNEFCVXGLGTCDAAFLNRLNKGCKDNE  
 GANDINPATEETPRANCPCSSPKICENAGIFTGIRDBEMCRVYCGALIKOR  
 MGTCTEDVHVSQKNAFEBGDLDAECNAGIFTGIRDBEMCRVYCGALIKOR  
 NGALINDOQIILVRLALRWVEYFLEDYFKIKKLPCTIBNGSTCINGNCKNRY  
 GEMTILKQDEWTKIKNHYLEKNGEDKNVTSLVNLTETVQLAAANDKQSTLCK  
 LKTSIGCNCPSNRKNDENDALIDCMANKLETYHIECTKHESVNSDQPHNCG  
 NPPBEDLLEENPEVQPGCFPTQOEPDCKGKLEKCKEKEQEPQABEG  
 GAIVSPGSGSEADKGPVPAIIPKQEPDLSGAVIISLSTLAWVGCFGA  
 FTYFLAKTKTSSVGNLFOILLPKSDYDIPFLTSPNRYLYTSGLKRGKXYILG  
 SGTDSYTDHNTSDITSSSEBYENDINDIYVPGSPKTYLLEVLARSGNNTAGN  
 TTASGNTTASGKNTPSQNDIQNDGIPSSKCTDNEKQKDSITSOYLOSENPT  
 NMLGYNNVNTHTPTSSHNVKEKPIKSIHORNPSGEVETIRGSEANFINDST  
 NSMSDLSNNHSPYNDLVSGIDLINDALSNNHIDLYDMLEKKEKNEHGLTGHK  
 HTNTYNAKPPARDPITNQINDLFKMLDRHNDCEKMNNEHRLPKLEKNEHPS  
 DINGSLPSGNVLTAVSIOIDMDPKTKNEITNMOTWPDSTMDTILDLLEKNEPS  
 YDPEDEDIITHDVVEKSSMDIYVDNNVTNNNMVFTQTHIEMNIVNNKEIFER

LEENPVKAPNI CPKVEPEPVYEEKCDLAPAKSEKSETEENSGEGSNEQNSKPK  
 EEPBPPTSSETPPAPPTIOPGQADQPNISDILASSTIPRGIALALTSIVLPL  
 KKKTSSVDLLAVLANIKPKEGMPKTKSNRYIPVASBNGKGTIYMGDSGHHY  
 EDTDVTSSSESEYSLDINDIYVPGSPKTYLLEVLBPSGNNTAGSGNTPSDND  
 IONDGIPSSKITDNEWNTLKDEFISQYLOSQPNVNDIYTSNGSTNTNTITTSRHN  
 VEKPIKSIHORNLYTGEINYNVNMVYMDPIPIRNDNYSIGDILINDLNDGYD  
 IYDEVLRKENELEGTINHYKQTSISVAKPARDPIHQQLLEFKMLDRHMDCKELK  
 NHEERLAKLEKEMNETHSNTHPSDNTLATVDSIOIQMDNPRPINFETMDPTILD  
 DLDINNEPYVDQDDIYVDVNDHSTVDSNAMDVSKVQIEMDVNTKLVEKXYPIRD  
 VMDI"

EXON  
 EYPSIDITWNT"  
 <11749..18339  
 /gene="var-2"  
 /note="putative"  
 /number=1  
 18340..19337  
 /gene="var-2"  
 /gene="var-2"  
 19338..>20738  
 /gene="var-2"  
 /note="putative"  
 /number=2

## ORIGIN

Query Match 9.5%; Score 109.2; DB 3; Length 22243;  
 Best Local Similarity 49.1%; Pred. No. 3.3e-08;  
 Matches 288; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

EXON  
 59 TACAAAACAATTTATGTTTCATTGAGATTAGAGAGGTAAGAGAGAGAGAGAGATTAA 118  
 18617 TTAAG 18676  
 INTRON  
 119 AAAAAAGTCTTATCTCTTGTCTGTAATTAATTAATTAAGAGCTTAACCTTTAT 178  
 18677 AAAAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18736  
 EXON  
 179 ATTAATATTTGTAATGAGTTTCTAGTCATGACACCACTCAGACAGAGATTTCAGAA 238  
 18737 ATTAATATTTATTAATTAATTAATTAATTTGATAGATTAATTAATTAATTAATTAAT 18796  
 EXON  
 239 AACAAATTTGTTAAACATCTTAATTAAGAACTTTAGTTAGCTTGAAGTTAGATTAA 298  
 18797 AAAAAATTTAAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18856  
 EXON  
 299 CAAAAAATTTACACGAGAGAAACCAATTAACCACTACCGTACAGTTATCATAGAT 358  
 18857 AGAAG 18916  
 EXON  
 359 GAATGTTTGTATATCATTAATTAATTAACACACAGAGAGAGAGAGAGAGAGAGAGAG 418  
 18917 AAAAAATTTAAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18976  
 EXON  
 419 ATGTTTATCATTAATTTTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 478  
 18977 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19036  
 EXON  
 479 AGAGTTATGATATCAATATTTTAATTAATTAATTTTAATTAATTAATTTTAATTAAT 538  
 19037 AATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19096  
 EXON  
 539 TTGAGATCTCTGACACTGTGTGATCATTATGATGATGATGATGATGATGATGATGAT 598  
 19097 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19156  
 EXON  
 599 TCAACAGAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 644  
 19157 TTTACATATCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19202

## RESULT 12

PVDLBC1 1395 bp DNA linear PLN 30-MAR-1995  
 LOCUS  
 DEFINITION Phaseolus vulgaris dlec1 gene for phytohemagglutinin (PHA-B)  
 (cystathionine synthase phytohemagglutinin).

ACCESSION X02408  
 VERSION X02408.1 GI:20985  
 KEYWORDS glycoprotein; haemagglutinin; lectin; phytohemagglutinin; signal peptide.

SOURCE  
 ORGANISM Phaseolus vulgaris  
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosoid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Phaseolus.

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT	FEATURES
		1 (bases 1 to 1395)					source
	Hoffman, U.M. and Donaldson, D.D.	Characterization of two Phaseolus vulgaris phytohemagglutinin genes					
		Characterization of two Phaseolus vulgaris phytohemagglutinin genes					
		EMBO J. 4 (4), 883-889 (1985)					
		85257505					
		2990911					
		Data kindly reviewed (06-MAR-1986) by D.D. Donaldson.					
		location/Qualifiers					
		1..1395					





6749 6848: gap of 100 bp  
6849 8101: contig of 1253 bp in length  
8102 8201: gap of 100 bp  
8202 9503: contig of 1302 bp in length  
9504 9603: gap of 100 bp  
9604 10826: contig of 1223 bp in length  
10827 10925: gap of 100 bp  
10926 12171: contig of 1245 bp in length  
12172 12271: gap of 100 bp  
12272 13557: contig of 1286 bp in length  
13558 13657: gap of 100 bp  
13658 14957: contig of 1300 bp in length  
14958 15057: gap of 100 bp  
15058 16339: contig of 1282 bp in length  
16340 16439: gap of 100 bp  
16440 17747: contig of 1308 bp in length  
17748 17847: gap of 100 bp  
17848 19125: contig of 1278 bp in length  
19126 19225: gap of 100 bp  
19226 20543: contig of 1318 bp in length  
20544 20643: gap of 100 bp  
20644 21933: contig of 1290 bp in length  
21934 22033: gap of 100 bp  
22034 23288: contig of 1255 bp in length  
23289 23388: gap of 100 bp  
23389 24637: contig of 1249 bp in length  
24638 24737: gap of 100 bp  
24738 26080: contig of 1343 bp in length  
26081 26180: gap of 100 bp  
26181 27463: contig of 1283 bp in length  
27464 27563: gap of 100 bp  
27564 28816: contig of 1253 bp in length  
28817 28916: gap of 100 bp  
28916 30228: contig of 1312 bp in length  
30229 30328: gap of 100 bp  
30329 31624: contig of 1296 bp in length  
31625 31724: gap of 100 bp  
31725 33059: contig of 1335 bp in length  
33060 33159: gap of 100 bp  
33160 34450: contig of 1291 bp in length  
34451 34550: gap of 100 bp  
34551 35813: contig of 1263 bp in length  
35814 35913: gap of 100 bp  
35914 37183: contig of 1270 bp in length  
37184 37283: gap of 100 bp  
37284 38505: contig of 1222 bp in length  
38506 38605: gap of 100 bp  
38606 39921: contig of 1316 bp in length  
39922 40021: gap of 100 bp  
40022 41265: contig of 1244 bp in length  
41266 41365: gap of 100 bp  
41366 42680: contig of 1315 bp in length  
42681 42780: gap of 100 bp  
42781 44112: contig of 1332 bp in length  
44113 44212: gap of 100 bp  
44213 45557: contig of 1345 bp in length  
45558 45657: gap of 100 bp  
45658 47001: contig of 1344 bp in length  
47002 47101: gap of 100 bp  
47102 48400: contig of 1299 bp in length  
48401 48500: gap of 100 bp  
48501 49780: contig of 1280 bp in length  
49781 49880: gap of 100 bp  
49881 51183: contig of 1303 bp in length  
51184 51283: gap of 100 bp  
51284 52519: contig of 1236 bp in length  
52520 52619: gap of 100 bp  
52620 53880: contig of 1261 bp in length  
53881 53980: gap of 100 bp  
53981 55274: contig of 1294 bp in length  
55275 55374: gap of 100 bp  
55375 56691: contig of 1317 bp in length  
56692 56791: gap of 100 bp

56792 58065: contig of 1274 bp in length  
58066 58165: gap of 100 bp  
58166 59539: contig of 1314 bp in length  
59540 59639: gap of 100 bp  
59640 60905: contig of 1266 bp in length  
60906 61005: gap of 100 bp  
61006 62302: contig of 1297 bp in length  
62303 62402: gap of 100 bp  
62403 62596: contig of 1194 bp in length  
62597 63896: gap of 100 bp  
63897 64984: contig of 1288 bp in length  
64985 65084: gap of 100 bp  
65085 66268: contig of 1184 bp in length  
66269 66368: gap of 100 bp  
66369 67636: contig of 1268 bp in length  
67637 67736: gap of 100 bp  
67737 69026: contig of 1290 bp in length  
69027 69126: gap of 100 bp  
69127 70399: contig of 1273 bp in length  
70400 70499: gap of 100 bp  
70500 71763: contig of 1284 bp in length  
71764 71863: gap of 100 bp  
71864 73116: contig of 1253 bp in length  
73117 73216: gap of 100 bp  
73217 74715: contig of 1439 bp in length  
74716 74815: gap of 100 bp  
74816 76106: contig of 1291 bp in length  
76107 76206: gap of 100 bp  
76207 77422: contig of 1216 bp in length  
77423 77522: gap of 100 bp  
77523 78820: contig of 1298 bp in length  
78821 78920: gap of 100 bp  
78921 80230: contig of 1310 bp in length  
80231 80330: gap of 100 bp  
80331 81610: contig of 1280 bp in length  
81611 81710: gap of 100 bp  
81711 83067: contig of 1357 bp in length  
83068 83167: gap of 100 bp  
83168 84434: contig of 1267 bp in length  
84435 84534: gap of 100 bp  
84535 85791: contig of 1237 bp in length  
85792 85891: gap of 100 bp  
85892 86978: contig of 1087 bp in length  
86979 87078: gap of 100 bp  
87079 88352: contig of 1274 bp in length  
88353 88452: gap of 100 bp  
88453 89751: contig of 1299 bp in length  
89752 91085: contig of 1234 bp in length  
91086 91185: gap of 100 bp  
91186 92443: contig of 1258 bp in length  
92444 92543: gap of 100 bp  
92544 93791: contig of 1248 bp in length.  
93791: Location/Qualifiers  
1. 93791  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"

FEATURES  
source

Query Match 8.3%; Score 95.2; DB 2; Length 93791;  
Best Local Similarity 35.4%; Pred. No. 4.3e-06;  
Matches 247; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

QY 12 AATGAAATGCTGTTTCTTCATCAGACAGAGGAGGACAGTTAACAACAAT 71  
DB 49600 ANNNANNN 49541  
QY 72 TTAATGTTCAATTTGAAGATTAGAGAGTAAAGAGAAAAAGATTAAAAATGCTCC 131  
DB 49540 NNN 49481  
QY 132 ATCTCTTGTCTTCTGTAATTAATTAATAGAGCTAAACCTTATATATATATGTA 191

Wed Jun 2 16:36:20 2004

us-10-015-637-1.rge

Page 15

[illegible]

Search completed: June 2, 2004, 04:50:37  
Job time : 4665 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:08:39 ; Search time 3381 Seconds  
(without alignments)  
10139.540 Million cell updates/sec

Title: US-10-015-637-1  
Perfect score: 1148  
Sequence: 1 taggattcctcaatagaaaa.....agagatgacgttaacga 1148

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: em\_estba:\*  
10: em\_estba:\*  
11: em\_estba:\*  
12: em\_estba:\*  
13: em\_estba:\*  
14: em\_estba:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_estba:\*  
18: em\_estba:\*  
19: em\_estba:\*  
20: em\_estba:\*  
21: em\_estba:\*  
22: em\_estba:\*  
23: em\_estba:\*  
24: em\_estba:\*  
25: em\_estba:\*  
26: em\_estba:\*  
27: em\_estba:\*  
28: em\_estba:\*  
29: em\_estba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	117.8	10.3	1101 29	CNS00EVL
2	104.6	9.1	1210 29	CG749728 P044-1-CO
3	103.2	9.0	1201 9	AL536104
4	101.8	8.9	1201 9	AL565455

C 5	99.6	8.7	1191 29	CG754863
C 6	99.4	8.7	1200 13	CG754863
C 7	98.8	8.5	1201 13	CG754863
C 8	97.4	8.5	1201 13	CG754863
C 9	97.4	8.5	1193 29	CG745316
C 10	97.4	8.5	1201 13	CG745316
C 11	96.8	8.4	1200 13	CG745316
C 12	96.6	8.4	1200 13	CG745316
C 13	96.6	8.4	1200 13	CG745316
C 14	96.6	8.4	1200 13	CG745316
C 15	95.8	8.3	1348 29	CG749499
C 16	94.6	8.2	932 9	AL514901
C 17	93.6	8.2	1201 13	CG745316
C 18	93.2	8.1	1201 29	CNS01JRG
C 19	93.2	8.1	1201 29	CNS01JRG
C 20	93.2	8.1	1201 29	CNS01JRG
C 21	92.8	8.1	1200 13	CG745316
C 22	92.4	8.0	1092 29	CNS020K7
C 23	92.4	8.0	1101 29	CNS020K7
C 24	92.4	8.0	1190 29	CNS020K7
C 25	91.8	8.0	1099 13	CG754863
C 26	91.4	8.0	1048 13	CG754863
C 27	91.4	8.0	1201 13	CG754863
C 28	91.4	8.0	1201 13	CG754863
C 29	91.4	8.0	1201 13	CG754863
C 30	91.2	7.9	1175 29	CG750201
C 31	91.2	7.9	1392 29	CG750201
C 32	90.6	7.9	1466 14	CG750201
C 33	90.6	7.9	1201 13	CG750201
C 34	90.4	7.9	1201 13	CG750201
C 35	90.4	7.9	1201 13	CG750201
C 36	90.4	7.9	1201 13	CG750201
C 37	90.2	7.9	1135 29	CG750201
C 38	90.2	7.8	1566 29	CG750201
C 39	89.8	7.8	1566 29	CG750201
C 40	89.6	7.8	1095 28	CG750201
C 41	89.6	7.8	1201 13	CG750201
C 42	89.6	7.8	1364 29	CG750201
C 43	89.4	7.8	895 29	CG750201
C 44	89.4	7.8	1098 13	CG750201
C 45	89.4	7.8	1200 13	CG750201

## ALIGNMENTS

RESULT 1  
LOCUS CNS00EVL  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29823 of RPL1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL069706.1 GI:494849  
VERSION AL069706  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Bukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygora;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (07-JUN-1999) Genoscope - Centre National de Sequenage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuo Oseawa and  
Aaron Mammeter in Pieter de Jong's laboratory in the Department of



QY 493 TACAAATATTTAAAAATATATTTAAAAACATTCCTAAAGTCATTCAGATCTCTCA 532  
 DB 321 NANN 262  
 QY 553 CACCTGTGTGATCTTTAGTCTGTATGTAGTACATCATCTGTACTTGTGACAGAGATTA 612  
 DB 261 AA 202  
 QY 613 AATAAATAGATTAAGGAT 672  
 DB 201 AA 142  
 QY 673 ATCAATTTAGAAAT 686  
 DB 141 AAAAAAAAAAAATTT 128

RESULT 3  
 AL536104/c 1201 bp mRNA linear EST 31-MAY-2003  
 LOCUS AL536104 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 DEFINITION CS0DF022YC18 5-PRIME, mRNA sequence.

ACCESSION AL536104.2 GI:31260974  
 VERSION AL536104  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12939597.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DF022BB09QPL.

FEATURES  
 source  
 1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF022YC18"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo (dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."

## ORIGIN

Query Match 9.0%; Score 103.2; DB 9; Length 1201;  
 Best Local Similarity 39.7%; Pred. No. 2.3e-08;  
 Matches 173; Conservative 85; Mismatches 171; Indels 7; Gaps 1;

QY 89 TTAAGGAGGTAAAG 148  
 DB 1085 YTAAGGAGGTAAAG 1026  
 QY 149 ATATAATATATAG 208  
 DB 1025 WAAAGGAGGTAAAG 966  
 QY 209 GAGCAGCAGTCAG 268

DB 965 ATTAATGCTAT 906  
 QY 269 TTTAGTAAAGCTCTGAG 328  
 DB 905 TATATAATTAAG 846  
 QY 329 AACCCAGTCAG 388  
 DB 845 AA 789  
 QY 389 AACCAAAATACATCTATATATATATATATATATATATATATATATATATATAT 448  
 DB 788 -----WATTAATTAATATATATATATATATATATATATATATATATATATAT 733  
 QY 449 GAGTTTTCAG 508  
 DB 732 AATTAATTAAT 673  
 QY 509 ATATATATTTAAAAAA 524  
 DB 672 HAHTTATTAATTAATTA 657

RESULT 4  
 AL565455 1201 bp mRNA linear EST 12-MAY-2003  
 LOCUS AL565455 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 DEFINITION CS0DF005Y018 3-PRIME, mRNA sequence.

ACCESSION AL565455  
 VERSION AL565455.2 GI:30549492  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 16, 2001 this sequence version replaced gi:12916848.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 9232.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DF005BH09NP1&cluster=9232.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DF005BH09NP1.

FEATURES  
 source  
 1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF005Y018"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo (dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."

## ORIGIN

Query Match 8.9%; Score 101.8; DB 9; Length 1201;  
 Best Local Similarity 37.0%; Pred. No. 4.1e-08;  
 Matches 228; Conservative 115; Mismatches 287; Indels 3; Gaps 3;

QY 53 AACAGTTAAACAAACAAATTTATGTTTCAATTTGAGATTAGAGAGGTAAAGAGAGAGAG 112

[illegible]

```

source
1..1391
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/lib="Epp EcorI BAC Library"
/notes="The library was generated by a partial digest of the genomic DNA with EcorI and cloning into the BAC vector."

ORIGIN
Query Match      8.7%; Score 99.6; DB 29; Length 1391;
Best Local Similarity 48.7%; Pred. No. 9,3e-08;
Matches 329; Conservative 0; Mismatches 339; Indels 8; Gaps 3;

   12 AATTGAAAAATGTGTAATTTCCATCACCGAGCAAAAGGGCAACGTTAACMAACAAT 71
    Db AAATTANTAAATATTTTTTAAATAAATAATTAATAATTAATTAATTAATTAATTAAT 956
       72 TTAGTTTCATTTGAGATTAAAGAGGTAAAGAGAAGAAAAGATTAAAAAAATGCTT 131
    Db TAAATATATATATATATAATAAATATVAAATTAATAAAAAAAAAAAATTAATATAT 896
   Qy 132 A---TCCTCTTGTTCTCGTAATTAATATATAGAGACTTAACTTTAAATTAATNTG 188
    Db 895 AATTATATAAAAATTAATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 836
   Qy 189 TAATTAGTTTTCAGTCATGAGCACCCTCAGAGCAAGATTCAAGAAACAATTTTG 248
    Db 835 ATATTAATAAATTAATATTAATATTAATTAATTAATTAATTAATTAATTAATTTT 776
   Qy 249 TTAACAATCTTATTAGAACTTTTGTGAATCTTGAAGTTAGCATTAACMAAAAAAT 308
    Db 775 TTATATTAATATATATATATATATATATATATATATATATATATATATATATTAAT 716
   Qy 309 TAGCAGAGAAACACATTAACCCACGACGCGAGGTTATCATAGAGATAATGTTT 368
    Db 715 TATTATAAAAAATTAATATTAATTAATTTTNTATTAATTAATTAATTAATTAATTAAT 656
   Qy 369 GATATCATTAATTAACACACACMAAATAATCATTAATTAACAAATATGTTATACA 428
    Db 655 AAAA-AAATTAATAAATAATATATTAATAAATAATATATATATATATATATATAT 597
   Qy 429 TATATTTTGTAAAACTTAGAGTTTTCAAAAACATCTAATTAACATTAATGTTATTA 488
    Db 596 TTTATATTAATTAATTAATTTTAAATTTT---ATTATATATTAATTAATTAATTAAT 541
   Qy 489 GAATATCAAAATATTAAAAAATTAATTTTAAAAAACAATCTAAAGCATTCAGATCT 548
    Db 540 TAAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 481
   Qy 549 CTCACACCTGTGCATCAATTGATCATGTATGTACATCATCTGTATGTTCAACA 608
    Db 480 TAAAAATTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 421
   Qy 609 GTAAATTAATAATTAAGATACTAGAGATATATATTAATTAATTAATTAATTAATTA 668
    Db 420 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 361
   Qy 669 GAAATCAAAATTAGAA 684
    Db 360 AAAAAAAAAAAAAAAAANNA 345

RESULT 6
BX415878/c 1200 bp mRNA linear EST 15-MAY-2002
LOCUS BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP0081104
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX415878
VERSION BX415878.1 GI:30765550
KEYWORDS EST.
SOURCE Homo sapiens (human)

```





ORIGIN vector "

Query Match 8.5%; Score 97.4; DB 29; Length 1193;  
Best Local Similarity 46.5%; Pred. No. 2.4e-07;  
Matches 313; Conservative 0; Mismatches 357; Indels 3; Gaps 2;

43 ACAAAGGCGACAGTTACAAACAAATTTATGTTTATTTGATTTGAGATTAAGAGGTTAG 102  
1024 AAAAAATTTTAAAAATTTAAAAATTTAAATTTTAAATTTAAAAATTTAAATTTAA 965  
103 GAGAAAAAAGATTAA-AAAAAATGCTCTTATCTCTTCTTCTGTAATTAATTAATTAAG 161  
964 ATATATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 905  
162 AGACTTAACTTTTAAATTAATTAATTTAAATTTAGTTTCTAGTCAGCACTCAG 221  
904 ATTAATTTATTAATTTAAATTTAAAAATTTAAATTTAAATTTAAATTTAAATTTAA 845  
222 AGACAAAGTTTCAAGAAACAAATTTGTTAAACATCTTATTAGAACTTTAGTTAGTC 281  
844 TATATAAATTTTAAAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTA 785  
282 TTGAAGTTAGATTTAAACAAAAAATTTACACGACGAGAACACATTAACCCACTACCT 341  
784 TAAATTTATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 725  
342 CAGGTTATCAATAGAGTGAATGTTTGTATTCATTAATTAATTAACACGACGAAATCA 401  
724 ATTAAAAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 665  
402 TCTAATTTAAACAAATTTATGTTATCATATATTTTGTAAAACTTTAGTTTCAAAA 461  
664 TAAATTTATTAATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 605  
462 CATCTTAATCAATGATTTAGAGTTTACAGAAATCAAAATTTTAAATTTAAATTTTAA 521  
604 ATNAAAAATTTAATTTTAA-ATAATTAATTTAAATTTAAATTTAAATTTAAATTTAA 547  
522 AAAACATCTTAAAGCATTCAGATCTCTCACACCTGTGTGATCACTTATGATCATGTAT 581  
546 ATAAATTTATTTATTTAAATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAA 487  
582 AGTACAAATCATTTGATTCACACAGAGTAAATTAATTAATTAATTAATTAATTTAA 641  
486 ATTAATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 427  
642 ATAAATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 701  
426 AAAAAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 367  
702 GACACAACTCACCC 714  
366 NNNNNNNNNCCCC 354

RESULT 10  
BX462207/c 1201 bp mRNA linear EST 22-MAY-2003  
LOCUS BX462207 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
DEFINITION clone CS0D004YB03 5-PRIME, mRNA sequence.  
ACCESSION BX462207.1 GI:31023422  
VERSION BX462207.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polajnar, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 24.r For more  
information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D004CA020p1&cluster=24.r>. Contact :  
Peng Liang Email: [liang@life.com](mailto:liang@life.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D004CA020p1.

## FEATURES

source  
1..1201  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D004YB03"  
/tissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo (dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."

## ORIGIN

Query Match 8.4%; Score 97; DB 13; Length 1201;  
Best Local Similarity 42.6%; Pred. No. 2.8e-07;  
Matches 253; Conservative 50; Mismatches 287; Indels 4; Gaps 1;

100 AAGGAAAGAAAAAGATTTAAAAAATGCTTTATCTCTTGTTCGTGATTAATATATA 159  
1135 AAAAAAATTTAAAAAATTTAAAAATTTAAAAATTTAAAAATTTAAAAATTTAAAA 1076  
160 AGAGACTTAACTTTTAAATTAATTTAAATTTAGTTTCTAGTCAGACCACTC 219  
1075 AAAAAAATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1016  
220 AGACAAAGTTTCAAGAAACAAATTTGTTAAACATCTTATTAAGAACTTTAGTAA 279  
1015 AAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 957  
280 TCTTAAGTTTGAATTTAAACAAAAAATTTACACGAGAACCAATTAACCACTACC 339  
956 ---AAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAA 900  
340 GTCAAGTTTATCAAGAGTGAATGTTTGTATTCATTAATTAATTAATTAATTAATTA 399  
899 TTTTAAAAAATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 840  
400 CATCTAATTTTAAACAAATTTGTTATCATATTTTGTAAAAAATTTAGAGTTTCA 459  
839 AAAAAAATTTAAAAAATTTAAAAAATTTAAATTAATTAATTAATTTAAATTTTAA 780  
460 AACATCTTAATTAATTAATTTAGATTTAAGAAATTAATTAATTTAAATTAATTTAA 519  
779 AAAAAATTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAATTTT 720  
520 AAAAAAATTTTAAAGCATTCAGATCTCTCACACCTGTGTGATCACTTATGATCATGTAT 579  
719 TAAAAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 660  
580 GTAGTACATCATTTGATTCACACAGAGTAAATTAATTAATTAATTAATTAATTAATTA 639  
659 TAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 600  
640 ATATTAATTTATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTT 693  
599 AAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTT 546

RESULT 11  
BX437758









```
Oy 457 CAAAACATTCATACATGATTAGAGTTATAGAAATACAAATATTTAAAAATTAATT 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 AAAAAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 365
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 517 TTAATAAAACATTTTAAAGTCATTCAGATCCTCTCACACCTGTGATCAATTAAGTCATG 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 TAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 577 TATGTAGTACATTCATGTAGTTCACACAGTAAATTAATTAAGATTAACCTAGGAA 636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 AAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 637 TATATTAATATATCAATTAATAATAATAATAATAATAATAATAATAATA 684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 ATTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: June 2, 2004, 03:32:47  
Job time : 3393 secs

Wed Jun 2 16:36:20 2004

us-10-015-637-1.rml

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 01:03:24 ; Search time 108 Seconds  
(without alignments)  
5898.922 Million cell updates/sec

Title: US-10-015-637-1

Sequence: 1 tagagatcctcaatagaagaa.....agagagtcgagtgtaatgca 1148

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued Patents NA:\*  
2: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	6.4	19124	2	US-08-487-826B-13
2	72.2	6.3	7218	1	US-08-232-463-14
3	72	6.3	5562	4	US-10-204-708-63
4	71.8	6.3	6866	4	US-10-204-708-63
5	70.6	6.1	837	3	US-08-998-416-288
6	68.2	5.9	6669	3	US-10-204-708-6
7	67	5.8	636	3	US-08-998-416-1137
8	66.6	5.8	8093	4	US-10-204-708-32
9	65.8	5.7	10467	4	US-10-204-708-2
10	65.8	5.7	19124	2	US-08-487-826B-13
11	64.8	5.6	6306	4	US-10-204-708-50
12	64.6	5.6	6317	4	US-10-204-708-12
13	64	5.6	5610	4	US-10-204-708-54
14	63.8	5.6	660	1	US-07-991-867B-32
15	63.8	5.6	660	1	US-08-107-755A-32
16	63.8	5.6	660	2	US-08-544-332-32
17	63.8	5.6	660	4	US-09-370-861A-32
18	63.8	5.6	1511	1	US-07-991-867B-8
19	63.8	5.6	1511	1	US-08-107-755A-8
20	63.8	5.6	1511	2	US-08-544-332-8
21	63.8	5.6	1511	2	US-09-370-861A-8
22	63.8	5.6	4810	4	US-08-852-629-11
23	63.8	5.6	4838	3	US-08-852-629-15
24	63.8	5.6	6040	4	US-10-204-708-69
25	62.6	5.5	6124	4	US-08-213-419B-3
26	62.4	5.4	6156	4	US-10-204-708-60
27	62.2	5.4	8961	4	US-10-204-708-80

c 28	61.4	5.3	640681	4	US-09-790-988-1	Sequence 1, Appl
c 29	61.2	5.3	6317	4	US-10-204-708-11	Sequence 11, Appl
c 30	60.6	5.3	665	2	US-08-883-795A-36	Sequence 36, Appl
c 31	60.2	5.2	1501	4	US-09-457-037B-30	Sequence 30, Appl
c 32	60.2	5.2	1501	4	US-09-733-151-30	Sequence 30, Appl
c 33	59.8	5.2	6801	4	US-10-204-708-61	Sequence 61, Appl
c 34	59.4	5.2	694	4	US-09-457-037B-39	Sequence 39, Appl
c 35	59.4	5.2	694	4	US-09-733-151-39	Sequence 39, Appl
c 36	59.4	5.2	1279	4	US-09-457-037B-40	Sequence 40, Appl
c 37	59.4	5.2	1279	4	US-09-733-151-40	Sequence 40, Appl
c 38	59.4	5.2	640681	4	US-09-790-988-1	Sequence 1, Appl
c 39	59.2	5.2	5610	4	US-10-204-708-53	Sequence 53, Appl
c 40	59.2	5.2	6124	4	US-08-213-419B-3	Sequence 3, Appl
c 41	59	5.1	11049	4	US-10-204-708-24	Sequence 24, Appl
c 42	59	5.1	53332	4	US-09-801-861-3	Sequence 3, Appl
c 43	58.6	5.1	615	3	US-08-998-416-186	Sequence 186, App
c 44	58.6	5.1	6113	4	US-10-204-708-13	Sequence 13, Appl
c 45	58.2	5.1	1218	2	US-08-731-722-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 593827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhan  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH21.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-0176  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 6.4%; Score 74; DB 2; Length 19124;  
Best Local Similarity 48.3%; Pred. No. 2.8e-07;

Wed Jun 2 16:36:20 2004

us-10-015-637-1.rn1

Page 2

Matches 304; Conservative 0; Mismatches 315; Indels 11; Gaps 3;

```
QY 41 AGACAAAGGGGACAGCTTACAAAACAAATTATGTTTCCTTTAGATTAAGAGAGTA 100
DB 15490 AATTTAAAAAAGAAAAAAGAAAAAAGAAAAAATTTTAAAAAATATATATATA 15549
QY 101 AGGAAAGAAAAAGATTAAAAAATGCTTATCTCTTGTCTGTATATATATA 160
DB 15550 ATTTAAATATATAATTTGATGATATATAAATGAAAAAGATTATCAAAAAAATTA 15609
QY 161 GAGACTTAACCTTTATATATATATATATATATATATATATATATATATATAT 220
DB 15610 AATATTTTATATATATATATATATATATATATATATATATATATATATAT 15666
QY 221 GAGACAGATTTCAAGAAAAACAATTTTGTAAACATCTTATTAAGAACTTTAGT 280
DB 15667 AAAAAACATTTAAAAAATATATATATATATATATATATATATATATATATAT 15726
QY 281 CTGGAAGTATGATTAAGAAAAAATTAACAGAGAAACAAATTAACCCCTAGT 340
DB 15727 ATTTAAATATATATATATATATATATATATATATATATATATATATATAT 15786
QY 341 TCAGGTTATCATAGAGTAAATGTTTGTATATATATATATATATATATATATAT 400
DB 15787 TATATCATTAATAAATATATATATATATATATATATATATATATATATATAT 15846
QY 401 ATCTAATTAACATATATATATATATATATATATATATATATATATATATATAT 460
DB 15847 AATTAATATATATATATATATATATATATATATATATATATATATATATAT 15903
QY 461 ACATTTATATATATATATATATATATATATATATATATATATATATATATAT 520
DB 15904 AAAAAATTAATGAAAAAATATATATATATATATATATATATATATATATAT 15963
QY 521 AAAAAATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 580
DB 15964 ATTAATATATATATATATATATATATATATATATATATATATATATATAT 16023
QY 581 TAGTACA-----ATCATTTGATGATGATGATGATGATGATGATGATGATGATG 635
DB 16024 ATTTACATATATATATATATATATATATATATATATATATATATATATAT 16083
QY 636 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 665
DB 16084 ATATATATATATATATATATATATATATATATATATATATATATATATAT 16113
```

RESULT 2  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.  
APPLICANT: SCHEFFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOMULOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313  
FILING DATE: 26-AUG-1991  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)836-4109  
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONING: pTZ19-FL8

US-08-232-463-14

Query Match 6.3%; Score 72.2; DB 1; Length 7218;  
Best Local Similarity 6.6%; Pred. No. 5.1e-07;  
Matches 23; Conservative 204; Mismatches 122; Indels 0; Gaps 0;

```
QY 685 TTTTATATATATATATATATATATATATATATATATATATATATATATATAT 744
DB 1127 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 1186
QY 745 CCACACATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCT 804
DB 1187 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 1246
QY 805 CATGTTCCACATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCT 864
DB 1247 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 1306
QY 865 CACCTACGCTCCACCTCTCTCATATATGAGCTATGAGCTATGAGCTATGAGCT 924
DB 1307 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 1366
QY 925 CACTCTCTCACTCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984
DB 1367 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 1426
QY 985 AATTTAACTATATATATATATATATATATATATATATATATATATATATATAT 1033
DB 1427 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 1475
```

RESULT 3  
US-10-204-708-63/c  
Sequence 63, Application US/10204708  
Patent No. 6677731

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 63  
LENGTH: 5562  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-63

Query Match 6.3%; Score 72; DB 4; Length 5562;  
Best Local Similarity 47.4%; Pred. No. 5.2e-07;  
Matches 216; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

QY 253 ACATCTTATFAGAACTTTAGTTAGTCTTGAAGTTAGATTAAACAAAAAATTACA 312  
DB 1864 ACTTCAATTAATAATAATACTAAACCTAATAATAATAATAATAATAATAATA 1805  
QY 313 CACGAGAAACATTAACCCCTACCGTACGTTATCATAGATGAATGTTGATA 372  
DB 1804 AAAAAAATCTACCATCTAATAATAATAATAATAATAATAATAATAATAATA 1745  
QY 373 TCATTAATATFACACACACACAAATACATCTAATATATACATATATATACATA 432  
DB 1744 AATTCATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1685  
QY 433 TTTTGTAAACCTTAGAGTTTTCAAACATTCATTAATACATATAGATTATAGAA 492  
DB 1684 ATATATATATTAATAATATATATATATATATATATATATATATATATATAT 1625  
QY 493 TACGAATTTTAAATAATATATTTTAAATAATAATAATAATAATAATAATAATA 552  
DB 1624 AAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1565  
QY 553 CACCTGTGATCTTTAGTCATGTATGTATGATCAATCATCTAGTTTACACAGATTA 612  
DB 1564 AATCATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1505  
QY 613 AATTAATAGGATTAATACATGATATATATATATATATATATATATATATAT 672  
DB 1504 AATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1445  
QY 673 ATCAATTAAGATTTTGTATTCCTCATGACATGACAA 708  
DB 1444 ATCAATAATAATAATCATCTTCAACGAAAAA 1409

RESULT 4  
US-10-204-708-20/c  
Sequence 20, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PILSENBRICK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 20

LENGTH: 6866  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-20

Query Match 6.3%; Score 71.8; DB 4; Length 6866;  
Best Local Similarity 48.0%; Pred. No. 5.2e-07;  
Matches 205; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 108 AAAAAAGTTAAAAAATGCTTATCTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 167  
DB 1146 AAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1089  
QY 168 AATCTTTAT 227  
DB 1088 AAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1029  
QY 228 GATTCAAGAAACATTTTGTAAACATCTTATATAGAACTTTAGTTAGTTGAAAG 287  
DB 1028 AATTCGAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 969  
QY 288 TTAGATTAACAAAAAATTTACACGAGAAACATTAACCCATCAAGTCAAGT 347  
DB 968 AATTTCTAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 909  
QY 348 ATCATAGAGTGAATGTTTGTATATCTTAATATATATATATATATATATATATAT 407  
DB 908 CTAAAT 849  
QY 408 TTAAACAT 467  
DB 848 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 789  
QY 468 AATTCATGTTAGTTTATATATATATATATATATATATATATATATATATATAT 527  
DB 788 AAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 729  
QY 528 TTCTAAA 534  
DB 728 ATAAAA 722

RESULT 5  
US-08-998-416-288/c  
Sequence 288, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippson, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII  
FILE REFERENCE: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwalis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416



APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
NUMBER OF INVENTION: AND USES THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwalis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: Pf/S-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1137:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: PAG1692RP  
ORGANISM: PAG1692RP  
US-08-998-416-1137

Query Match 5.8%; Score 67; DB 3; Length 636;  
Best Local Similarity 44.9%; Pred. No. 3.5e-06;  
Matches 253; Conservative 0; Mismatches 310; Indels 0; Gaps 0;

QY 130 TTATCTCTTTCTGTTAATAATAATAGAGACTTAACTTTTAAATAATAATTTGT 189  
DB 636 TTTTATAAGTATTTTAACTACATCTTTATAATAATTTTAAATAATAATGAT 577

QY 190 AATTAGGTTTCTAGTATGAGCACCCTGAGACAGATTTCAAGAAACATTTTGT 249  
DB 576 AAAATATTAATTAATAATTAATTAATAATAATAATAATAATAATAATTAAT 517

QY 250 TAAACATCTTTATTAGAACTTTTATAGTCTTGAAGTTTGAATTTAAACAAAAAAT 309  
DB 516 TATTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 457

QY 310 ACACAGAAACCAATAAACCCATACCTACCTGAGTTATCATGAAGATGTTTGT 369  
DB 456 TTCTTATAAAGATTTAAATTAATAATAATAATAATAATAATAATAATAATAATA 397

QY 370 ATATCATTAATAATAACACACAAAAATACATCTAATTAATAATAATAATAATAAT 429  
DB 396 ATAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 337

QY 430 ATATTTTGTAAAACTTTAGAGTTTTCAAAAACATCTTAATACATGATTTAG 489

DB 336 ATTTTAATAACAAATTAATAATAATAATAATAATAATAATAATAATAATAAT 277  
QY 490 AAATACAAATATTAAAAAATAATAATAATAATAATAATAATAATAATAATAAT 549  
DB 276 TAAAGAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 217  
QY 550 TCACACCTGCTGATCATTTAGTATGATGATGATGATGATGATGATGATGATGAT 609  
DB 216 TAAATAGTATTCATATAATAATAATAATAATAATAATAATAATAATAATAAT 157  
QY 610 TAAAAATAAAGGATAAAGGATAAAGGATAAAGGATAAAGGATAAAGGATAAAGG 669  
DB 156 TAAAGTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 97  
QY 670 AAATCAAAATTAAGATTTTGTAT 692  
DB 96 ATAATGATAATAATAAGTTTAAAT 74

RESULT 8  
US-10-204-708-32/c  
Sequence 32, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
TITLE OF INVENTION: by Assessing DNA Methylation  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
PRIOR FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 32  
LENGTH: 8093  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-32

Query Match 5.8%; Score 66.6; DB 4; Length 8093;  
Best Local Similarity 46.2%; Pred. No. 9.5e-06;  
Matches 222; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY 197 TTTTCTAGTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 256  
DB 2357 TTTTAAATCATCTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2298

QY 257 CTTATTAGAACTTTTGTAGTCTTGTAGTCTTGTAGTCTTGTAGTCTTGTAGTCT 316  
DB 2297 AATTAAACATATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2238

QY 317 AGAAACACATAAACCCATCTACCTGAGTTATCATGAAGTCAATTTTGTATCAT 376  
DB 2237 CAATAACACCTTAAAGCTTATATCAATTTTATTAATCAATTTTATTAACGTCC 2178

QY 377 TAAATATAACACACAAAAATACATCTAATTAATAAATAATAATAATAATAATTT 436  
DB 2177 GTAGTTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2118

QY 437 TGTAAACCTTACAGTTTTCAAAAACATCTTAATACATGATTTAGAGTTTATAG 496





US-08-487-826B-13

Query Match 5.7%; Score 65.8; DB 2; Length 19124;  
Best Local Similarity 44.8%; Pred. No. 1.9e-05;  
Matches 291; Conservative 0; Mismatches 357; Indels 1; Gaps 1;  
QY 50 AACAAACAAATTTATGCTCTTCTGTTCTGTAATAATAATAGAGACTTTAAACCTTTTAATA 119  
DB 18279 AATTAAAGGAATAAAGCTTAATAATAATAATAAGAAATAGTATATTTTATTAATAATAACA 18220  
QY 120 AAAAATGTCCTTATCTCTTCTGTTCTGTAATAATAATAGAGACTTTAAACCTTTTAATA 179  
DB 18219 AGAAATATTTGTTATATTAATAATAATTTTATTTAATAGGAACTATATATATTTGTTAT 18160  
QY 180 TAAATATTTGTAATTTAGGTTTCTGTTCTGTAATAATAATAGAGACTTTAAACCTTTTAATA 239  
DB 18159 ATAATAATTTTATATGAGATTTATATTTTCTGTTCTGTAATAATAATAGAGACTTTTAATA 18100  
QY 240 ACAATTTTGTAAACATCTTATAGAACTTTTGTAGTTTAAAGCTTTGAGCTTGAAGT-TAGAATTA 298  
DB 18099 GTACAAAATAAACAATATATAAACAATATATAAACAATATATAAACAATTTTAAATAATAT 18040  
QY 299 CAAAAAAATTTACACAGAGAAACAAATATAACCACTACCGTCAGGTTATCATAGGAT 358  
DB 18039 TATAAATAATCAACAATAATAATAATAATAATACTGTAATAATAATAATAATAATAATA 17980  
QY 359 GAAATGTTTGTATATCATTTAATAATAACACACACACACACACACACACACACACACATAT 418  
DB 17979 TGAAGGATATATATTTAGAAAGAGCAATATATATATATATATATATATATATATATACG 17920  
QY 419 AAGTTATACATATATTTTGTGAAACCTTAGAGTTTTCACAAACATTTCTAATACATGAT 478  
DB 17919 AAAAAATGTACAGAAAGAAACAAATAAATCAATATATATATATATATATATATATATAT 17860  
QY 479 AGAGTTTATAGAAATACAAATATTTAAATAATAATTTTAAATAACATTTCTAAGTCA 538  
DB 17859 AGATACAAATTTATATCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 17800  
QY 539 TTCAGATCCTCTCACACCTGTGATCATTTTGTATGATGATGATGATGATGATGATGATGAT 598  
DB 17799 ATTACAAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 17740  
QY 599 TCACACAGAGTAAATAATAATAAGATTAACCTAGGATATATATATATATATATATATATAT 658  
DB 17739 TGATTTAAACAAACAAACATATTTTTCACAAATATACCAAAACAAATTAATAATAATA 17680  
QY 659 AAAAAAGGGAATCAATAGAAATTTTGTGTTTCCCAATGACACACACACACACACACACAC 707  
DB 17679 ATAAATAAATAAACAATATATACAAATAATAAACAATATATATATATATATATATAT 17631

## RESULT 11

US-10-204-708-50/c  
; Sequence 50, Application US/10204708  
; Patent No. 667731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 50  
; LENGTH: 6306  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-50

Query Match 5.6%; Score 64.9; DB 4; Length 6306;  
Best Local Similarity 46.3%; Pred. No. 2.2e-05;  
Matches 315; Conservative 0; Mismatches 362; Indels 3; Gaps 3;  
QY 10 TCATAGAAATGTTGTTTCTCTCATCCACGACAAAGGGCAACAGTTTAAACAAACAA 69  
DB 3409 TTAATAATAATCTTCTTAACCTACAAATAAACAACGACATTTAACTCTTAATCTCACAAA 3350  
QY 70 ATTATATGTTTCATTTGAGATTTAAGGAGTTAAGCAAGAAAAAGATTTAAAAAATGCTC 129  
DB 3349 ATACTCTTTAAAAAATAAATAAATCACTATTACAAATATCGCACCAAACTCACACCTAT 3290  
QY 130 TTATCTCTTTGTTTCTGTAATAATAATAATAAGAGACTTTAACTTTTAAATATATAATGT 189  
DB 3289 AATTCCAATATCTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3230  
QY 190 AATTAGGTTTCTAGTCATGAGCACTCTCAGAGACAGAGATTTCAAGAAAAACAAATTTGT 249  
DB 3229 AATAAACCTTAATCAATACATTAACGAAACCCCATCTTAACAAAAAATAAACAACAAACA 3170  
QY 250 TAAACATCTTTATAGAAACCTTTTGTAGTTTAAAGTTTGAAGTTAGAAATTA-AAACAAAAAAT 308  
DB 3169 AAAAACTTTTATTAACCTTAATAATAATAATAATAATAATAATAATAATAATAATAATA 3110  
QY 309 TACACAGGAAACACAAATAAACCCCTACCGTCAGGTTATCATAGGATGAAATGTTT 368  
DB 3109 TAAATAAATAAATAATCGGTTAACTCGAATTTAAATAAATAAATAAATAAATAAATAAATA 3050  
QY 369 GATATCATTAATAATATACACACACAAATAATACATCTTAATTTATAACAATA-TGTTATAC 427  
DB 3049 ACTACACTCCAACTCTAAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATA 2990  
QY 428 ATATATTTTGTAAAACTTAGAGTTTTCACAAACATTTCTAATACATGATTAGAGTTTAT 487  
DB 2989 ACATTAATATATATCTCTCCCTATATAATACATCAACTTACTTTTACAAACAATAC 2930  
QY 488 AGAAATACAAATTTTAAAAAATAATAATTTTAAAAAATAAATCTTAAAGTCATTCAGATCC 547  
DB 2929 TAAATAACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2870  
QY 548 TCTCACACCTGTGATCATTTAGTCATGATGATGATGATGATGATGATGATGATGATGATG 607  
DB 2869 AACCAACACTTTTATATACACCACTAACACACATTTCTAC-ATCTTTTAAATAAATCCTT 2811  
QY 608 AGTAAATAAATAAAGGATAAATAAGGATAATATATAATAATAATAATAATAATAATAATA 667  
DB 2810 TAAATAACCTGTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2751  
QY 668 GGAATAATCAAAATAGAAATTT 687  
DB 2750 AAAAACTATCATTTAATTTAT 2731

## RESULT 12

US-10-204-708-12/c  
; Sequence 12, Application US/10204708  
; Patent No. 667731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1





Query Match 5.6%; Score 63.8; DP 1; Length 660;  
Best Local Similarity 48.5%; Pred. No. 1.8e-05;  
Matches 272; Conservative 0; Mismatches 277; Indels 12; Gaps 3;

Qy	109	AAAGATTAAAGAAATGCTCTTATCTCTTTGTTCTGTAATAATAATAAGAGACTTA	168
Db	637	ATAAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	578
Qy	169	AACTTTAAATATAATAATAATAATAATAATAATAATAATAATAATAATAATA	228
Db	577	TACITTTTGGAAATTTATCAATGATATATATTTTAAACATTGAAGATATACTA	518
Qy	229	ATTCAAGAAACAAATTTTGTAAACATCTTATTAGAAACCTTTTAGTCTGTGAAGT	288
Db	517	ATTTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	460
Qy	289	TAGAAATAACAAAAAATTTACACAGAGAAACACAAATAAACCCACTACCGTCAGGTTA	348
Db	459	CAATTTAACTCACTACTACTATAGGAAACAACTATTAAAGTTACCAATTTATTTT	400
Qy	349	TCATAAGGATCAAAATGTTTTCATATCATTAATAATAACACACACAAATAACATCTAAT	408
Db	399	AGATATTATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	345
Qy	409	ATACCAATATATGTTATACATATATTTTGTAAAAAATTTAGAGTTTTCAAAAACATCTA	468
Db	344	ATATTACATGATTCACAAATTTAAAAATTTCTATAGATGTTGTTAGTATATATTTTATA	285
Qy	469	ATACATGTTAGAGTTTATAGAAATACAAATATTTTAAAAAATNTAATTTTAAAAAACA	528
Db	284	TGCTATTTTGTATATAAGATATATCTAAATATGTTATATTTTAAATTTTGTATATAA	225
Qy	529	TCTAAAGTCATTCAGATCCTCTCACACCTGTGTATCATTTTAGTCTATGTAGTA---	585
Db	224	TTAAATTAATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTCTGGAATTA	165
Qy	586	--CAATCATTTAGTTTCAACACAGAGTAAATAATAAGGATAAACTAGGGAATATATAT	643
Db	164	TTTAAATATTTATTCAT	105
Qy	644	AATATATCAATTAATAAAA	664
Db	104	AATATATTAATTAATAACA	84

Search completed: June 2, 2004, 02:26:17  
Job time : 112 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 01:00:19 ; Search time 538 Seconds  
(without alignments)  
9064.928 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148

Sequence: 1 taggaccccttcataagaaaa.....agagagtgatggttaatgca 1148

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002s:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	1148	6	ABN83916 Truncated
2	1148	100.0	1832	6	ABN83929 Arcelin-5
3	1144.8	99.7	1821	6	ABD39066 Phaseolus
4	1122	97.7	1122	6	ABN83917 Truncated
5	796.6	69.4	1866	6	ABN83928 Arcelin-4
6	795	69.3	1872	6	ABN83927 Arcelin-3
7	790	7.8	8056	7	ABZ10246 Haematopo
8	87	7.6	3683	7	ABZ10199 Haematopo
9	35.2	7.4	4846	6	ABK87142 Scarlet r
10	85	7.4	2888	6	ABN83926 Arcelin-5
11	85	7.4	8056	7	ABZ10100 Haematopo
12	84.8	7.4	8056	7	ABZ10246 Haematopo
13	84.4	7.4	1211	2	AAQ94051 Kidney be
14	83.4	7.3	5925	6	ABL33576 Human inm
15	83	7.2	26997	4	AA546748 Tumour su
16	82.6	7.2	1501	7	ABZ10188 Haematopo
17	82.6	7.2	1501	9	AD54162 Human lym
18	82.2	7.2	16786	6	ABL34157 Human inm
19	81.4	7.1	3683	7	ABZ10053 Haematopo
20	81.2	7.1	6216	6	ABK39932 Human che
21	81.2	7.1	6216	6	ABL70139 Chemical
22	80.6	7.0	14006	6	ABL33958 Human inm
23	80.2	7.0	6048	6	ABQ67002 Human ang

## ALIGNMENTS

### RESULT 1

ABN83916

ID ABN83916 standard; DNA; 1148 BP.

XX AC

XX ABN83916;

XX DT 06-SEP-2002 (first entry)

XX DE Truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.

XX KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;

XX KW pharmaceutical; ds.

XX OS Phaseolus vulgaris.

XX XX WO200250295-A2.

XX PD 27-JUN-2002.

XX PP 17-DEC-2001; 2001WO-US047495.

XX PR 18-DEC-2000; 2000US-0255879P.

XX (RENE-) RENESSEN LLC.

XX Wang Q, Dubois P, Liang J, Oulmassov T;

XX WPI; 2002-508809/54.

XX New transformed or transgenic soybeans plants or cells with an Arcelin-5

XX promoter, for use as an improved dietary source of protein for humans or

XX animals, or for producing soybeans with important agricultural or

XX nutritional properties.

XX Claim 4; Page 65-66; 74pp; English.

XX The invention relates to a transformed soybean plant cell and transgenic

XX soybean plant, both of which has a nucleic acid molecule comprising the

XX phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.

XX The transformed soybean plant cell and transgenic soybean plant are

XX useful as an improved source of dietary protein for humans and livestock.

XX These are also useful for producing soybean plants that exhibit important

XX agricultural, nutritional or pharmaceutical properties. The current

XX sequence represents the truncated P. vulgaris exotic genotype G02771

XX Arcelin-5 promoter sequence. This sequence is a deletion mutant of the

XX Arcelin-5 promoter, created by the removal of approximately 600 base

C 24	80	7.0	83391	6	ABQ67093	Abq67093 Human ang
C 25	79.8	7.0	8056	7	ABZ10100	Abz10100 Haematopo
C 26	79.8	7.0	21354	4	AA546815	AA546815 Tumour su
C 27	79.2	6.9	6644	2	AA333181	AA333181 Base sequ
C 28	79.2	6.9	6568	6	ABL33397	ABL33397 Human inm
C 29	79.2	6.9	7372	2	AA333182	AA333182 Base sequ
C 30	79.2	6.9	7797	2	AA333180	AA333180 Cowpox vi
C 31	79.2	6.9	7892	6	ABK40056	Abk40056 Human che
C 32	79.2	6.9	7996	2	AA333184	AA333184 Base sequ
C 33	79.2	6.9	14615	4	AA546704	AA546704 Tumour su
C 34	79	6.9	6106	4	AA546429	AA546429 Tumour su
C 35	79	6.9	6106	6	ABK40031	Abk40031 Human che
C 36	79	6.9	6106	6	ABL33472	ABL33472 Human inm
C 37	78.6	6.8	778	6	ABQ15588	Abq15588 Oligonucle
C 38	78.6	6.8	778	6	ABQ15589	Abq15589 Oligonucle
C 39	78.6	6.8	12237	6	ABL34358	ABL34358 Human inm
C 40	78.4	6.8	8693	6	ABL32808	ABL32808 Human inm
C 41	78.2	6.8	5654	4	AA546623	AA546623 Tumour su
C 42	78.2	6.8	5654	6	ABL33874	ABL33874 Human inm
C 43	78.2	6.8	6175	6	ABL33307	ABL33307 Human inm
C 44	77.8	6.8	15548	6	ABL34155	ABL34155 Human inm
C 45	77.8	6.8	18218	6	ABL33949	ABL33949 Human inm



[illegible]

RESULT 3	
AAD29066	
ID	AAD29066 standard; DNA; 1821 BP.
XX	
AC	AAD29066;
XX	
DT	07-MAY-2002 (first entry)
XX	
DE	Phaseolus vulgaris arcelin promoter.
XX	
KW	Heterologous gene expression; plant; arcelin promoter; arcelin leader;
KW	seed preferred expression cassette; ds.
XX	
OS	Phaseolus vulgaris.
XX	
PN	WO200200899-A2.
XX	
FD	03-JAN-2002.
XX	
PF	31-MAY-2001; 2001WO-EP006298.
XX	
PR	29-JUN-2000; 2000EP-00202278.
XX	
PA	(VLA-) VLAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX	
PI	Angenon G, De Jaeger G, Goossens A, Depicker A;
XX	
DR	WPI; 2002-139925/18.
XX	
PT	Novel seed preferred expression cassette having gene regulatory elements
PT	comprising arcelin promoter, arcelin 51 leader, and arcelin 51 3' end,
PT	useful for heterologous gene expression in plants.
XX	
PS	Claim 1; Page 46-47; 52pp; English.
XX	
CC	The invention relates to heterologous gene expression in plants. The
CC	invention also relates to seed preferred expression cassette having gene
CC	regulatory elements comprising arcelin promoter, arcelin 51 leader, and
CC	arcelin 51 3' end. This expression cassette is useful for heterologous
CC	gene expression in plants. The protein encoded by the heterologous gene
CC	is a single chain antibody variable fragment (scFv). The present sequence
CC	is Phaseolus vulgaris arcelin promoter of the invention
XX	
SQ	Sequence 1821 BP; 698 A; 308 C; 210 G; 605 T; 0 U; 0 Other;
	Query Match      99.7%; Score 1144.8; DB 6; Length 1821;
	Best Local Similarity 99.8%; Pred. No. 1.7e-189;
	Matches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 TAGATCCTTCATAGAAAATGTATTTCCTCATCACGACAAGGGGCAACAGTTA 60
DB	674 TAGATCCTTCATAGAAAATGTATTTCCTCATCACGACAAGGGGCAACAGTTA 733
QY	61 ACAAAACAAATTTATGTTCATTTGCAGTAAAGCAGGTGAAGGAAGAAAAAGATTAAA 120
DB	734 ACAAAACAAATTTATGTTCATTTGAGATTAAAGCAGGTGAAGGAAGAAAAAGATTAAA 793
QY	121 AAAATGTCCTTACTCTTGTTCTGTGTAATAATAATAAGACACTTTAACTTTTAAAT 180
DB	794 AAAATGTCCTTACTCTTGTTCTGTGTAATAATAATAAGACACTTTAACTTTTAAAT 853
QY	181 AATAATGTAAATTAGGTTTTCTAGTCATGAGCACCATCGAGAGACAGATTTCAGAAAA 240
DB	854 AATAATGTAAATTAGGTTTTCTAGTCATGAGCACCATCGAGAGACAGATTTCAGAAAA 913
QY	241 CAATTTGTTAAACATCTTATTAGAACTTTTAGTTAAGTCTTCAAGTTAGAAATTAACA 300
DB	914 CAATTTGTTAAACATCTTATTAGAACTTTTAGTTAAGTCTTCAAGTTAGAAATTAACA 973
QY	301 AAAAAAATTACACACGAGAAACRCAATAAACCCCACTACCGTCAGGTTATCATAGGATGA 360



974	AAAAAAGTACACACGCGAAACACAATAAACCCCACTACCGTCAAGTGTATCATTAAGGATGA	1033
361	AATGTTTTGATATCATTAATAATAACACACACAAAAAATACATCTAAATTAATAACAATATAT	420
1034	AATGTTTTGATATCATTAATAATAACACACACAAAAAATACATCTAAATTAATAACAATATAT	1093
421	GTATACATATATTTTTGTAAACACTTAGAGTTTTTCAAAACAATCTCTAAATACATGATTAG	480
1094	GTATACATATATTTTTGTAAACACTTAGAGTTTTTCAAAACAATCTCTAAATACATGATTAG	1153
481	AGTTTATAGAAATACAAATATTTTAAAAATATATTAATTTTAAAAAAAACATTTCTAAAGTCATT	540
1154	AGTTTATAGAAATACAAATATTTTAAAAATATATTAATTTTAAAAAAAACATTTCTAAAGTCATT	1213
541	CAGATCCTCTCCACACCTGTGTGATCATTTTAGTTCATGTATGTAGTACAAATCAATTTGTAGTTC	600
1214	CAGATCCTCTCCACACCTGTGTGATCATTTTAGTTCATGTATGTAGTACAAATCAATTTGTAGTTC	1273
601	ACAACAGAGTAAATAAATAAGGATAACTAGGGAATATATATATATATACAAATTAAT	660
1274	ACAACAGAGTAAATAAATAAGGATAACTAGGGAATATATATATATATACAAATTAAT	1333
661	AAAAAGGAGAAATCAAAATTAGAATTTTTTGATTTCCCACTAGACAGAACTCACCATGCGAC	720
1334	AAAAAGGAGAAATCAAAATTAGAATTTTTTGATTTCCCACTAGACAGAACTCACCATGCGAC	1393
721	GCTGGCACTTCAGCTGCTGCTCTCTCCACACATGCTCTCATGCTCACTTTGGACTTTGGCTTT	780
1394	GCTGGCACTTCAGCTGCTGCTCTCTCCACACATGCTCTCATGCTCACTTTGGACTTTGGCTTT	1453
781	TTCACTATGACACAACTCGCCATGCAATTTGGCCACGTGAGCTCCTTCTCTTCCCATGAT	840
1454	TTCACTATGACACAACTCGCCATGCAATTTGGCCACGTGAGCTCCTTCTCTTCCCATGAT	1513
841	GCACCACTGGGATGCACTGCTGCCACCTCAGCTGCCACCTCTCTCATATTAGCGCTAC	900
1514	GCACCACTGGGATGCACTGCTGCCACCTCAGCTGCCACCTCTCTCATATTAGCGCTAC	1573
901	TGGCCATGCACACTGGCCACCTCAGCACTCCTCTCACTTCCCATGTCTACCTGCCAAACCG	960
1574	TGGCCATGCACACTGGCCACCTCAGCACTCCTCTCACTTCCCATGTCTACCTGCCAAACCG	1633
961	CTTCTCTCCATAAATATCTATTTAAATTTTAAACTTAAATTTTCAATACATTTTTTTTGATGA	1020
1634	CTTCTCTCCATAAATATCTATTTAAATTTTAAACTTAAATTTTCAATACATTTTTTTTGATGA	1693
1021	CGTGGATGCAATGCCATCGTTGTTTTTAATAATTTGTAATTTGGAGTTGGAATAATAAATGA	1080
1694	CGTGGATGCAATGCCATCGTTGTTTTTAATAATTTGTAATTTGGAGTTGGAATAATAAATGA	1753
1081	AAGAAAAAGCTTGGAAAGATTTTTCATTTGTTGTTATATAAATAGAAAGAGAGTGTGATGG	1140
1754	AAGAAAAAGCTTGGAAAGATTTTTCATTTGTTGTTATATAAATAGAAAGAGAGTGTGATGG	1813
1141	TTAATGCA 1148	
1814	TTAATGCA 1821	

RESULT 4  
ABN83917  
ID ABN83917 standard; DNA: 1122 BP.

AC ABN83917;

XX DT 06-SEP-2002 (first entry)

XX Truncated P. vulcaris exotic genotype G02771 Arcelin-5 promoter #2.

XX *Phaseolus vulgaris* var. *minutifolius* (L.) Fernald  
XX  
XX Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;  
KW pharmaceutical; ds.  
KW  
XX  
XX  
OS *Phaseolus vulgaris*.

XX	WO200250295-A2.	
PN		
XX	27-JUN-2002.	
PD		
XX		
XX	17-DEC-2001; 2001WO-US047495.	
PF		
XX		
XX	18-DEC-2000; 2000US-0255879P.	
PR		
XX		
XX	(RENE-) RENESSEN LLC.	
PA		
XX		
XX	Wang Q, Dubois P, Liang J, Oulmassov T;	
PI		
XX		
XX	WPI; 2002-508809/54.	
DR		
XX		
PT	New transformed or transgenic soybeans plants or cells with an Arcelin-5	
PT	promoter, for use as an improved dietary source of protein for humans or	
PT	animals, or for producing soybeans with important agricultural or	
PT	nutritional properties.	
XX		
XX	Example 5; Page 66; 74pp; English.	
PS		
XX		
XX	The invention relates to a transformed soybean plant cell and transgenic	
CC	soybean plant, both of which has a nucleic acid molecule comprising the	
CC	Phasolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.	
CC	The transformed soybean plant cell and transgenic soybean plant are	
CC	useful as an improved source of dietary protein for humans and livestock.	
CC	These are also useful for producing soybean plants that exhibit important	
CC	agricultural, nutritional or pharmaceutical properties. The current	
CC	sequence represents a truncated P. vulgaris exotic genotype G02771	
CC	Arcelin-5 promoter sequence. This sequence was used in the expression of	
CC	a GUS reporter gene in a soybean cotyledon	
XX		
SQ	Sequence 1122 BP; 410 A; 210 C; 146 G; 356 T; 0 U; 0 Other;	

```
Query Match          97.7%; Score 1122; DB 6; Length 1122;
Best Local Similarity 100.0%; Pred. No. 1.4e-185;
Matches 1122: Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

Qy	3	GGATCCTTCAATAGAAAATGTTGTTATTTCTCATCAACAGCAAGGGCGAACAGTTAAAC	52
Db	1	GGATCCTTCAATAGAAAATGTTGTTATTTCTCATCAACAGCAAGGGCGAACAGTTAAAC	60
Qy	63	AAAAACAAATTTATGTTTTCATTTGAGATTAAAGGAAGTAAAGGAAGAAAAGATTAAAAAA	122
Db	61	AAAAACAAATTTATGTTTTCATTTGAGATTAAAGGAAGTAAAGGAAGAAAAGATTAAAAAA	120
Qy	123	AATGTGCTTATCTCTTTGTTTCTGTATATATATATAGAGACTTAAACTTTTAAATATAA	182
Db	121	AATGTGCTTATCTCTTTGTTTCTGTATATATATATAGAGACTTAAACTTTTAAATATAA	180
Qy	183	TAATTGTGAATTAGTGTCTTAGTCATGAGCACCACCTCAGAGACAAGATTTCAAGAAACA	242
Db	181	TAATTGTGAATTAGTGTCTTAGTCATGAGCACCACCTCAGAGACAAGATTTCAAGAAACA	240
Qy	243	ATTTTGTTTAAACATCTTTATTAGAAAACCTTTAGTTTAAAGTCTTTGAAGTTAGAAATTAACAAA	302
Db	241	ATTTTGTTTAAACATCTTTATTAGAAAACCTTTAGTTTAAAGTCTTTGAAGTTAGAAATTAACAAA	300
Qy	303	AAAAATPACACACGAGAAAACAAATAAACCCACCTACCGTCAGGTGTATCATGAAGTGAAC	362
Db	301	AAAAATPACACACGAGAAAACAAATAAACCCACCTACCGTCAGGTGTATCATGAAGTGAAC	360
Qy	363	TGTTTTGATATCATTTAAATATAACACACAAAAATATCATCTAAATATATAACAATATATGT	422
Db	361	TGTTTTGATATCATTTAAATATAACACACAAAAATATCATCTAAATATATAACAATATATGT	420
Qy	423	TATACATATATTTTGTAAAAACCTTAGAGTTTTTCAAAACANTTCTAATACATGATTAGAG	482
Db	421	TATACATATATTTTGTAAAAACCTTAGAGTTTTTCAAAACANTTCTAATACATGATTAGAG	480
Qy	483	TTTATAGAAATACAAATATTTTAAAAAATAAATTTTAAAAAACAATTTCTAAAGTCATTCA	542







KW Human; haematopoietic cell proliferation disorder; cytostatic;  
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KW cytosine methylation state; gene; ds.

OS Homo sapiens.

PN WO20027272-A2.

XX 03-OCT-2002.

PD 26-MAR-2002; 2002WO-EP003401.

PF 26-MAR-2001; 2001US-0278333P.

PR (EPIC-) EPIGENOMICS AG.

PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;

PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;

PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;

PI Schwobe I, Ziebarth H;

XX WPI; 2003-018942/01.

DR Detecting and differentiating between hematopoietic cell proliferative  
 KW disorders, comprises contacting a target nucleic acid with a reagent that  
 KW distinguishes between methylated and non-methylated CpG dinucleotides.

PS Claim 28; SEQ ID NO 339; 117pp; English.

XX The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. AB29861 to AB21118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used; for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC lymphocytic leukaemia cells; for differentiating between acute  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related DNA  
 CC sequences. The nucleotide sequences from the present invention can also  
 CC be used for detecting a predisposition to, differentiation between  
 CC subclases, diagnosis, prognosis, treatment and/or monitoring of  
 CC haematopoietic cell proliferative disorders. The present method enables a  
 CC highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients

XX SQ Sequence 3683 BP; 716 A; 0 C; 760 G; 2207 T; 0 U; 0 Other;

Query Match 7.6%; Score 87; DB 7; Length 3683;

Best Local Similarity 46.6%; Pred. No. 4.4e-06;

Matches 312; Conservative 0; Mismatches 355; Indels 2; Gaps 1;

QY 54 ACAGTTAAACAAACAAATTTATGTTTCATTTGAGATTAGGAGCTAAGGAGAGAAAAAG 113

DB 3311 AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3252

QY 114 ATTTAAAAAATGCTCTTATCTCTTTGTTTCGTGTAATATATATAGAGACTTAAACTT 173

DB 3251 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3192

QY 174 TTAATATAATAATGTAATAGGTTTCTTCATGACCCACCTCAGAGCAAGATTTC 233

DB 3191 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3134

QY 234 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 293

DB 3133 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3074

QY 294 TTAACAAAAAAATTTACACACGAGAAACACAATAAACCCACTACCTCAGGTTATCAT 353  
 DB 3073 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3014  
 QY 354 AGGATGAATGTTTTCATATATCAATTAATAAACAACAAAAATACATCTAATTATAAC 413  
 DB 3013 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2954  
 QY 414 AATATATGTTATACATATATTTTGTAAAAAATCTTAGAGTTTTCAAAAACATCTAATAACA 473  
 DB 2953 TAAACAAAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2894  
 QY 474 TGATTAGAGTTTATAGAAATACAAATATTTAAAAAATAAATAAATAAATAAATAA 533  
 DB 2893 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2834  
 QY 534 AGTCATTAGATCTCTTCACACCTGCTGTGATCATTTTAGTGTATGTAGTACATCAT 593  
 DB 2833 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2774  
 QY 594 GTAGTTCACACACGAGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 653  
 DB 2773 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2714  
 QY 654 ATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 713  
 DB 2713 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2654  
 QY 714 CATGCACGC 722  
 DB 2653 CTTTCCAC 2645

RESULT 9

ID ABK87142 standard; DNA; 4846 BP.

AC ABK87142;

DT 07-OCT-2002 (first entry)

DE Scarlet runner bean C541 genomic sequence.

XX Expression cassette; promoter activity; suspensor cell; plant embryo;  
 KW modulation of gene transcription; Scarlet runner bean; C541; plant; gene;  
 KW transgenic; ds.

OS Phaseolus coccineus.

PK Key Location/Qualifiers

FT CDS 3134..3552

FT /\*tag= a  
 FT /product= "C541 protein"

PN WO200244333-A2.

PD 06-JUN-2002.

XX 28-NOV-2001; 2001WO-US044737.

XX 28-NOV-2000; 2000US-00724857.

PR 28-NOV-2000; 2000US-0235672P.

XX (REGC ) UNIV CALIFORNIA.

XX (CERS-) CERES INC.

XX Meterings K, Apuya NR, Tatarinova T, Goldberg RB;

PI WPI; 2002-508506/54.

DR P-PSDB; RAU98691.

XX Expression cassette comprises promoters with basal promoter activity

PT operably linked to a heterologous polynucleotide, useful for expression  
PT genes in suspensor cells in plants and/or basal region of plant embryo.  
XX  
PS  
XX Claim 25; Fig 4; 114pp; English.  
XX  
XX The present invention relates to expression cassettes comprising a  
CC promoter sequence and a promoter polynucleotide with basal promoter  
CC activity, where the promoter sequence is operably linked to a  
CC heterologous polynucleotide, and when the expression cassette is inserted  
CC into a plant, the heterologous polynucleotide is specifically expressed  
CC in a suspensor cell and/or basal region of a plant embryo. The invention  
CC also provides polynucleotide sequences encoding scarlet runner bean  
CC (Phaseolus coccineus) G564 and C541 proteins for use in the expression  
CC cassettes of the invention. The expression cassettes comprising promoters  
CC and promoter control elements are useful for modulating transcription of  
CC genes in a plant suspensor cell and/or basal region of a plant embryo.  
CC The present sequence represents Scarlet runner bean C541 genomic sequence  
XX  
SQ Sequence 4846 BP; 1894 A; 777 C; 753 G; 1415 T; 0 U; 7 Other;  
Query Match 7.4%; Score 85.2; DB 6; Length 4846;  
Best Local Similarity 65.5%; Pred. No. 9.2e-06;  
Matches 156; Conservative 0; Mismatches 78; Indels 4; Gaps 2;  
QY 464 TTCTATACATGATTACAGTTTATAGAAATACAAATATTTAAAAATATATTTAAAAA 523  
DB 4086 TTATATTACAAATTTATGACTTATAGAAATACAAATATTTAAAAATATA-AGGTTCAAAC 4144  
QY 524 AACATCTTAAGTCATTCAGATCCTCTCACACCTGTGTATCATTTAGTCATGTATAG 583  
DB 4145 TACATCTTAAGTCTTTCAGACCCCTCTGACACATGT---ATCATCTGCTGATATGTGA 4201  
QY 584 TACAATCATTTGTAGTTTCAACACAGAGTAAATATAAGGATATAAAGTACGGAATATATAT 643  
DB 4202 TACAGTCATCGCAGTTTCAACAGATACACAGAAACCAAGGTAAGCTAATGAAAAAAT 4261  
QY 644 AATATATACATTAATATAAAGGAAAGGAAATCAATCAATAGATTTTGGATCCCAT 701  
DB 4262 TCCATAACATATTAATTAATTCATGCAAAAGAACAGTCAAAAGTAATCATTTATAACAT 4319  
RESULT 10  
ABN83926  
ID ABN83926 standard; DNA; 288 BP.  
XX  
AC ABN83926;  
XX  
DT 06-SEP-2002 (first entry)  
XX  
DE Arcelin-5 promoter sequence fragment.  
XX  
KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;  
KW pharmaceutical; ds.  
XX  
OS Phaseolus vulgaris.  
XX  
PN WO200250295-A2.  
XX  
PD 27-JUN-2002.  
XX  
PF 17-DEC-2001; 2001WO-US047495.  
XX  
PR 18-DEC-2000; 2000US-0255879P.  
XX  
PA (RENE-) RENESSEN LLC.  
XX  
PI Wang Q, Dubois P, Liang J, Oulmassov T;  
XX  
DR WPI; 2002-508809/54.  
XX  
PT New transformed or transgenic soybeans plants or cells with an arcelin-5  
PT promoter, for use as an improved dietary source of protein for humans or  
PT animals, or for producing soybeans with important agricultural or

PT nutritional properties.  
XX  
PS Example 1; Fig 1; 74pp; English.  
XX  
XX The invention relates to a transformed soybean plant cell and transgenic  
CC soybean plant, both of which has a nucleic acid molecule comprising the  
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.  
CC The transformed soybean plant cell and transgenic soybean plant are  
CC useful as an improved source of dietary protein for humans and livestock.  
CC These are also useful for producing soybean plants that exhibit important  
CC agricultural, nutritional or pharmaceutical properties. The current  
CC sequence represents an arcelin-5 promoter sequence fragment  
XX  
SQ Sequence 288 BP; 121 A; 26 C; 38 G; 103 T; 0 U; 0 Other;  
Query Match 7.4%; Score 85; DB 6; Length 288;  
Best Local Similarity 100.0%; Pred. No. 7.9e-06;  
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1037 TCGTTGTTTAAATTAATTTGATTTGAGTTGAATTAATAATGAAAGAAAAAGTTGGAA 1096  
DB 204 TCGTTGTTTAAATTAATTTGATTTGAGTTGAATTAATAATGAAAGAAAAAGTTGGAA 263  
QY 1097 AGATTTGCAATTTGTTGTGTATAA 1121  
DB 264 AGATTTGCAATTTGTTGTGTATAA 288  
RESULT 11  
ABZ10100  
ID ABZ10100 standard; DNA; 8056 BP.  
XX  
AC ABZ10100;  
XX  
DT 16-JAN-2003 (first entry)  
XX  
DE Haematopoietic cell proliferation disorder related DNA sequence #240.  
XX  
KW Human; haematopoietic cell proliferation disorder; cytostatic;  
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
KW cytosine methylation state; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2002727272-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 26-MAR-2002; 2002WO-EP003401.  
XX  
PR 26-MAR-2001; 2001US-0278333P.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;  
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;  
PI Schwöpe I, Ziebarth H;  
XX  
DR WPI; 2003-018942/01.  
XX  
PT Detecting and differentiating between hematopoietic cell proliferative  
PT disorders, comprises contacting a target nucleic acid with a reagent that  
PT distinguishes between methylated and non-methylated CpG dinucleotides.  
XX  
PS Claim 28; SEQ ID NO 240; 117pp; English.  
XX  
XX The present invention describes a method for detecting and  
CC differentiating between haematopoietic cell proliferative disorders  
CC associated with at least 1 gene and/or their regulatory regions in a  
CC subject. The method comprises contacting a target nucleic acid in a  
CC biological sample obtained from the subject with at least 1 reagent,  
CC which distinguishes between methylated and non-methylated CpG



Db 2040 AAAAAATATATTTTATTTTAAATATAATTTAAATTTTATTTAAATTTTAAATATAAATAATTAAT 1981

Qy 310 ACACACAGAGAACACACATAAACCCACTACCGTCAGGTTATCATACAGGATCAAGATGTTTGG 369

Db 1980 ATTAATATTTTATTTTAAATAATTAATTTTATTTTATTTATATATTAATTAATTAATTTT 1921

Qy 370 ATATCATTAATAATACACACACAAAATAACATCTAATATATAACATATATATATATATATAT 429

Db 1920 TTTTATTAATAATTTTATTTTATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATA 1861

Qy 430 ATATTTTGTGTAACACTAGAGTTTTCAAAACATTTCTAATATATATATATATATATATATAT 486

Db 1860 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1801

Qy 487 TAGAATATACAAATATTTTAAATAATAATAATTTTAAATAAATAAATAAATAAATAAATAAATA 546

Db 1800 TAATACATATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1741

Qy 547 CTCTCACACCTGTGTGATCATTTAGTCAATGTATGTATGTATGTATGTATGTATGTATGTAT 606

Db 1740 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1681

Qy 607 GAGTAAATAATAAGGATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 662

Db 1680 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1625

RESULT 13

AAQ94051

ID AAQ94051 standard; DNA; 1211 BP.

XX

AC AAQ94051;

XX

DT 10-MAY-1996 (first entry)

XX

DE Kidney bean lectin-like gene.

XX

KW lectin-like; kidney bean; insect resistance; genomic analysis;

KN DNA marker; ss.

XX

OS Phaseolus vulgaris.

XX

PN JP07132092-A.

XX

PD 23-MAY-1995.

XX

PF 11-NOV-1993; 93JP-00305988.

XX

PR 11-NOV-1993; 93JP-00305988.

XX

PA (NORQ ) NORINSUISANSHO NOGYO SEIBUTSU SHIGEN.

XX

DR WPI; 1995-220115/29.

XX

PT New kidney bean gene encoding lectin-like protein - for the production of

PT insect-resistant crop plants.

XX

PS Claim 2; Fig 1-3; 6pp; Japanese.

XX

CC The DNA encodes a lectin-like protein isolated from kidney bean var.

CC Kentucky Wonder. The gene provides insect-resistance to crop plants, and

CC is also useful as a DNA marker in gene and genomic analysis

XX

SQ Sequence 1211 BP; 344 A; 322 C; 203 G; 342 T; 0 U; 0 Other;

Query March 7.4%; Score 84.4; DB 2; Length 1211;

Best Local Similarity 67.1%; Pred. No. 1.1e-05;

Matches 200; Conservative 0; Mismatches 31; Indels 67; Gaps 3;

Qy 917 CACCTCAGCACTCTCTCTCACTTCCCATTCGCAACACCGCTCTCTCCCAATAATA 976

Db 7 CACATGTGCTCTCTCTCACTTCTCACTGCTAAGTG-CACCCCGCTCTCTCCCAATAATA 65

Qy 977 TCTATTTAAATTTAACTAATTAATTTTCAATAT-----ACTTTTGTGATGAGTGGATGC 1029

Db 65 TCTCTTCACTTTAAACTAATTAATTTTCAATATTTTTCATATTTTTCATGATGACGTGGATGC 125

Qy 1030 ATTGCATCGTTGTTTAAATAATTTTAAATTTG----- 1061

Db 126 ATTGCAATCGTTGCTTAATCTTATTTTATATTTCTTCTCTCCCTCAATAATATTAC 185

Qy 1062 -----GAGTTGAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1090

Db 186 AAAAAAGAAAAAGTTAATCAATTCGAAAAACACGTTTAAATAACAACCAAGAAAAAG 245

Qy 1091 TTGGAAAGATTTTGCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1148

Db 246 ATCGAAGTTTTTTCAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 303

RESULT 14

ABL33576/c

ID ABL33576 standard; DNA; 5925 BP.

XX

AC ABL33576;

XX

DT 26-MAR-2002 (first entry)

XX

DE Human immune system associated gene SEQ ID NO: 1549.

XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antianaemic; cytosolic; neutropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;

XX

OS Homo sapiens.

XX

PN WO200200928-A2.

XX

PD 03-JAN-2002.

XX

PF 02-JUL-2001; 2001WO-BP007537.

XX

PR 30-JUN-2000; 2000DE-01032529.

XX

PA (EPIG-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K;

XX

DR WPI; 2002-130909/17.

XX

PT Nucleic acid comprising fragment of chemically modified gene, useful for

PT diagnosis and treatment of diseases associated with abnormal cytosine

PT methylation.

XX

PS Claim 1; SEQ ID NO 1549; 32pp + Sequence Listing; German.

XX

CC The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention

XX

SQ Sequence 5925 BP; 1510 A; 105 C; 1246 G; 3064 T; 0 U; 0 Other;

Query March 7.3%; Score 83.4; DB 6; Length 5925;

Best Local Similarity 49.5%; Pred. No. 1.9e-05;

Matches 351; Conservative 0; Mismatches 351; Indels 7; Gaps 5;





```
Db 5675 TAAATAACACATTTCTCRAAAAAAATAACAAATACCAACACATATAAAAAACACTCA 6616
Qy 425 TACATATATTTTGTAAAACTTAGAGTTTTCAAAAACATTTCTAAATACATGATTAGAGTT 484
Db 6615 ACATCACTTATTATCAAAAAAATAACAATCAAAACCCACAAATATAACCAATTTTACTCCT 6556
Qy 485 TATAGAAATACAAATATTTTAAAAAATATAATTTTAAAAAACATTTCTAAAAGTCATTGAGA 544
Db 6555 ACAAATAAACCTCATTTAAATAATAAAAAAATAATTAATTAACATATAATATAATAA 6496
Qy 545 TCCTCTCACACCTGTGTGATCATTTAGTCAATGTATGTAGTACAAATCATTTGTAGTTACAAA 604
Db 6495 AAAAAACACTTTTACACTTACCATAAAAAATATAAACTAAATACACCACTATAAAAAACAA 6436
Qy 605 CAGAGTAAATATAATAGGATTAACCTAGGGAATATATATATATATACAAATTAATAAAA 664
Db 6435 CATAAAAACTAATAACAACCACTATAAACTAATACACCACTATAAAAAACATATAAAA 6376
Qy 665 AAGGGAATCAAAATTAGAAATTT 687
Db 6375 AATTAAAAATAAACTACCAATTT 6353
```

Search completed: June 2, 2004, 02:24:19  
Job time : 546 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 2, 2004, 02:26:24 ; Search time 4650 Seconds  
(without alignments)  
10700.503 Million cell updates/sec  
Title: US-10-015-637-1  
Perfect score: 1148  
Sequence: 1 taggacccctcaatagaaaa.....acagagtgtgttaatgca 1148

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 3470272 seqs, 21671516995 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database : GenEmbl:  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pi.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.man.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1148	100.0	1148	6	AX463281	AX463281 Sequence
2	1148	100.0	1832	6	AX463294	AX463294 Sequence
3	1122	97.7	1122	6	AX463282	AX463282 Sequence
4	1046	91.1	1821	6	AX343309	AX343309 Sequence
5	1046	91.1	3900	8	PVARSXK1	ZS0202 P. vulgaris
6	107	9.3	1866	6	AX463293	AX463293 Sequence
7	107	9.3	1872	6	AX463292	AX463292 Sequence
8	85	7.4	288	6	AX463291	AX463291 Sequence
9	69	6.0	2288	8	AF193029	AF193029 Phaseolus
10	65	5.7	1992	8	PVDFLSC2	X04659 P. vulgaris
11	49	4.3	1889	8	PVDFLSC2	J01261 P. vulgaris
12	48	4.2	4564	8	PVDFLSC1	M68913 Phaseolus v
13	46	4.0	1768	8	PVDFLSC1	X04660 P. vulgaris
14	42	3.7	1852	8	AF255723	AF255723 Phaseolus
15	41	3.6	1211	6	E09281	E09281 DNA encodin
16	41	3.6	1395	8	PVDFLSC1	X02408 Phaseolus v
17	41	3.6	1441	8	PVDFLSCA	K03288 P. vulgaris
18	27	2.4	6875	9	AC096359	AC096359 Homo sapi
19	27	2.4	126318	9	AC128689	AC128689 Homo sapi
20	27	2.4	153165	2	AC015539	AC015539 Homo sapi
21	27	2.4	171589	2	AC073567	AC073567 Homo sapi
22	27	2.4	175466	2	AC016419	AC016419 Homo sapi
23	27	2.4	204954	2	AC128028	AC128028 Rattus no
24	27	2.4	276343	2	AC125570	AC125570 Rattus no
25	26	2.3	316	6	AX463290	AX463290 Sequence
26	26	2.3	322	6	AX463289	AX463289 Sequence
27	26	2.3	1188	8	PVDFLSC2A	X06336 P. vulgaris
28	26	2.3	1661	8	PVDFLSC2	X02409 Phaseolus v
29	26	2.3	1784	8	PVDFLSCB	K03289 P. vulgaris
30	24	2.1	32818	3	AC024825	AC024825 Caenorhab
31	24	2.1	148761	9	AL450352	AL450352 Human DNA
32	24	2.1	150601	9	AC117426	AC117426 Homo sapi
33	24	2.1	160099	2	AC073207	AC073207 Homo sapi
34	24	2.1	174383	5	AL329216	AL329216 Zebrafish
35	24	2.1	184119	9	AC090348	AC090348 Homo sapi
36	24	2.1	185796	2	AP001398	AP001398 Homo sapi
37	24	2.1	187023	2	AC006882	AC006882 Caenorhab
38	24	2.1	193595	2	AC073681	AC073681 Mus muscu
39	24	2.1	234679	2	EX511263	EX511263 Danio rer
40	24	2.1	251579	2	AC073677	AC073677 Mus muscu
41	23	2.0	13529	3	AY159038	AY159038 Dictyoste
42	23	2.0	40589	3	CEZK1025	AL022288 Caenorhab
43	23	2.0	40589	3	CEZK1025	AL022288 Caenorhab
44	23	2.0	46054	2	AC145619	AC145619 Homo sapi
45	23	2.0	66806	9	AP000022	AP000022 Homo sapi

ALIGNMENTS

RESULT 1  
AX463281  
LOCUS AX463281 1148 bp DNA linear PAT 15-JUL-2002  
DEFINITION Sequence 1 from Patent WO0250295.  
ACCESSION AX463281  
VERSION AX463281.1 GI:21886232  
KEYWORDS  
SOURCE Phaseolus vulgaris  
ORGANISM Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE 1  
AUTHORS Culmassov, T., Wang, Q., Dubois, P. and Liang, J.



959 AAAAAAATTACACACAGAAACAAATAAACCCACTACCGTCAGGTATCATCBAAGATGA 1018  
361 AATGTTTGGATATCATTAATAATAACACACACAAAAATACATCTAATATTAACAATATAT 420  
1019 AATGTTTGGATATCATTAATAATAACACACACAAAAATACATCTAATATTAACAATATAT 1078  
421 GTTATACATATATTTTGTAAACACTTAGAGTTTTCACAAACATCTTAATACATGATTAG 480  
1079 GTTATACATATATTTTGTAAACACTTAGAGTTTTCACAAACATCTTAATACATGATTAG 1138  
481 AGTTTATAGAAATACAAATATTTTAAATAATATTAATTTTAAAAAACAATCTTAAGTCAT 540  
1139 AGTTTATAGAAATACAAATATTTTAAATAATATTAATTTTAAAAAACAATCTTAAGTCAT 1198  
541 CAGATCTCTCACACCTGTGTGATCATTTAGTCATGTATGTAGTACATCATTTGTAGTTC 600  
1199 CAGATCTCTCACACCTGTGTGATCATTTAGTCATGTATGTAGTACATCATTTGTAGTTC 1258  
601 ACAACAGATGAAATAAATAAGGATAAAGTAACTAGGGAATATATATAATATATACAAATTAAT 660  
1259 ACAACAGATGAAATAAATAAGGATAAAGTAACTAGGGAATATATATAATATATACAAATTAAT 1318  
661 AAAAAAGGAAATCAAAATTTAGAAATTTTGAATCCCAACATGACACAACTCAGCATGCAC 720  
1319 AAAAAAGGAAATCAAAATTTAGAAATTTTGAATCCCAACATGACACAACTCAGCATGCAC 1378  
721 GCTGCCACCTCAGCTCCCTCTCTCCACACATGCTCATGTCATCTTCGACTTTGGCTTT 780  
1379 GCTGCCACCTCAGCTCCCTCTCTCCACACATGCTCATGTCATCTTCGACTTTGGCTTT 1438  
781 TTCACTATGACACAACTCGCCATGATGTTGCCACGTCAGTCTCTCTCTCTCTCTCTCTCT 840  
1439 TTCACTATGACACAACTCGCCATGATGTTGCCACGTCAGTCTCTCTCTCTCTCTCTCTCT 1498  
841 GACACACTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
1499 GACACACTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1558  
901 TGCCATGTCACACTGCCACCTCAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
1559 TGCCATGTCACACTGCCACCTCAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1618  
961 CTCTCTCTCAAT 1020  
1619 CTCTCTCTCAAT 1678  
1021 CGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1080  
1679 CGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1738  
1081 AAAAAAAGTTGGAAGATTTGCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140  
1739 AAAAAAAGTTGGAAGATTTGCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1798  
1141 TTAATGCA 1148  
1799 TTAATGCA 1806

RESULT 3  
AX463282  
LOCUS AX463282 1122 bp DNA linear PAT 15-JUL-2002  
DEFINITION Sequence 2 from Patent WO0250295.  
ACCESSION AX463282  
VERSION AX463282.1 GI:2186233

KEYWORDS  
SOURCE Phaseolus vulgaris  
ORGANISM Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE 1

AUTHORS Oulmasov, T., Wang, Q., Dubois, P. and Liang, J.  
TITLES Arcelin-5 promoter and uses thereof  
JOURNAL Patent: WO 0250295-A 2 27-JUN-2002;  
RENESSEN LLC (US)  
FEATURES Location/Qualifiers  
source 1..1122  
/organism="Phaseolus vulgaris"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3885"  
ORIGIN  
Query Match 97.7%; Score 1122; DB 6; Length 1122;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 GGATCCTTCAATAGAAAATGCTTATTTCTCATCACCAGACAAAGGGGCAACAGTTAAC 62  
DB 1 GGATCCTTCAATAGAAAATGCTTATTTCTCATCACCAGACAAAGGGGCAACAGTTAAC 60  
QY 63 AAAAAAATTTATGTTTCTATTTGAGATTAAAGGAGGTAAAGGAAAAAAGATTAAAAA 122  
DB 61 AAAAAAATTTATGTTTCTATTTGAGATTAAAGGAGGTAAAGGAAAAAAGATTAAAAA 120  
QY 123 AATGCTCTTATCTCTTGTGTTCTGTAATAATAATATAGAGACTTAACTTTTAAATAA 182  
DB 121 AATGCTCTTATCTCTTGTGTTCTGTAATAATAATATAGAGACTTAACTTTTAAATAA 180  
QY 183 TAATGCTTAATAGGTTTCTTAGTCATGAGCACCACCTCAGAGACAAAGATTTCAGAAAA 242  
DB 181 TAATGCTTAATAGGTTTCTTAGTCATGAGCACCACCTCAGAGACAAAGATTTCAGAAAA 240  
QY 243 ATTGTTTAAACATCTTATTTAGAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAACAAA 302  
DB 241 ATTGTTTAAACATCTTATTTAGAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAACAAA 300  
QY 303 AAAAAATTACACAGACGAAACACAAATAAACCCACTACCGTCAGGTATCATTAAGATGAAA 362  
DB 301 AAAAAATTACACAGACGAAACACAAATAAACCCACTACCGTCAGGTATCATTAAGATGAAA 360  
QY 363 TGTTTTGATATCATTAATAATAACACACACAAATAATACATCTAATATTAACAATATATGT 422  
DB 361 TGTTTTGATATCATTAATAATAACACACACAAATAATACATCTAATATTAACAATATATGT 420  
QY 423 TATACATATATTTTCTGTAAGCTTTAGGTTTTCACAAACATTTCTAATACATGATTAGAG 482  
DB 421 TATACATATATTTTCTGTAAGCTTTAGGTTTTCACAAACATTTCTAATACATGATTAGAG 480  
QY 483 TTTATAGAAATACAAATAATTTTAAAAAATAATAATTTTAAAAAACAATCTTAAAGTCATTCA 542  
DB 481 TTTATAGAAATACAAATAATTTTAAAAAATAATAATTTTAAAAAACAATCTTAAAGTCATTCA 540  
QY 543 GATCCTCTCACACCTGTGTGATCATTTAGTCATGTTAGTACATCAATCATTTAGTTTAC 602  
DB 541 GATCCTCTCACACCTGTGTGATCATTTAGTCATGTTAGTACATCAATCATTTAGTTTAC 600  
QY 603 AACAGAGTAAATAATAAATAAGGATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 662  
DB 601 AACAGAGTAAATAATAAATAAGGATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 660  
QY 663 AAAAGGAAATCAAAATTTAGAAATTTTGTATCCCACTGACACAACTGACCATGCAAC 722  
DB 661 AAAAGGAAATCAAAATTTAGAAATTTTGTATCCCACTGACACAACTGACCATGCAAC 720  
QY 723 TGCCACCTCAGCTCCCTCTCTCCACACATGTCATGTCACATTTTGGCTTTT 782  
DB 721 TGCACCTCAGCTCCCT 780  
QY 783 CACTATGACACAACTCGCCATGCAATGTCACAGTGAAGTCTCTCTCTCTCTCTCTCTCTCT 842  
DB 781 CACTATGACACAACTCGCCATGCAATGTCACAGTGAAGTCTCTCTCTCTCTCTCTCTCTCT 840  
QY 843 CACCATGCGCATGCTGTCACCTCAGTCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 902



MEDLINE 95045555  
PubMed 7957215  
REMARK (sites)  
REFERENCE 2  
AUTHORS Goossens A., Ardiles Diaz, W., De Keyser, A., Van Montagu, M. and Angenon, G.  
TITLE Nucleotide sequence of an arcelin5-I genomic clone from wild Phaseolus vulgaris  
JOURNAL Plant Physiol. 109, 722-722 (1995)  
REMARK (sites)  
REFERENCE 3  
AUTHORS Goossens A.  
TITLE Direct Submission  
JOURNAL Submitted (01-AUG-1995) Goossens A., Universiteit Gent, Laboratorium Genetika, K.L.Ledeganckstraat 35, Gent, Belgium, 9000  
REMARK Revised by [4]  
REFERENCE 4 (bases 1 to 3900)  
AUTHORS Goossens A.  
TITLE Direct Submission  
JOURNAL Submitted (13-AUG-1998) Goossens A., Universiteit Gent, Laboratorium Genetika, K.L.Ledeganckstraat 35, Gent, Belgium, 9000  
COMMENT On Aug 25, 1998 this sequence version replaced gi:929816.  
FEATURES  
source Location/Qualifiers  
1..3900  
/organism="Phaseolus vulgaris"  
/mol\_type="genomic DNA"  
/strain="G02771"  
/db\_xref="taxon:3885"  
/clone="GARCS-8A"  
/haplotype="arcs5"  
/tissue type="leaf"  
1707..1710  
CAAT\_signal  
/citation=[2]  
1790..1793  
TATA\_signal  
/citation=[2]  
1792..1797  
TATA\_signal  
gene  
1822..2754  
mRNA  
/gene="arcs5-I"  
1822..2754  
/gene="arcs5-I"  
/citation=[2]  
/evidence=experimental  
1822..1834  
5'UTR  
/gene="arcs5-I"  
/citation=[2]  
1835..2620  
gene  
/gene="arcs5-I"  
/citation=[1]  
1835..2620  
CDS  
/gene="arcs5-I"  
/citation=[2]  
/codon\_start=1  
/evidence=experimental  
/product="arcelin 5a"  
/protein\_id="CAA90585.1"  
/db\_xref="GI:3451282"  
/db\_xref="GOA:Q42460"  
/db\_xref="SWISS-PROT:Q42460"  
/translation="MSSKLLSLFLVLLTHNSATETSPNPNFHTDDKLILQNGA  
TISKGQLQGVGNSLPRVDSIGRAFYSDPIQKDSNVASNTNPTFIIRKNGS  
ISAYGLAFALVPVNSPPKKQKEFGIFNTNNPNPNTAVVNFNKRIDFDKNFIK  
PVYNENCDPHYNGEKTDVQIYDSSNNDLRFVLFHTVSVQKCVSATVHLEKVEDM  
VSVGFSAISGLTETDTTHDVLISWSFSKFRNKLNLNLIL"  
1835..1897  
sig\_peptide  
/gene="arcs5-I"  
/citation=[1]  
/evidence=experimental  
1898..2617  
mat\_peptide  
/gene="arcs5-I"  
/product="arcelin 5a"  
/citation=[1]

3'UTR /evidence=experimental  
2618..2754  
/gene="arcs5-I"  
/citation=[2]  
polyA\_signal /citation=[1]  
2704..2709  
/gene="arcs5-I"  
/citation=[1]  
polyA\_signal /gene="arcs5-I"  
2708..2713  
/gene="arcs5-I"  
polyA\_signal /gene="arcs5-I"  
2722..2727  
polyA\_signal /gene="arcs5-I"  
2726..2731  
ORIGIN  
Query Match 91.1%; Score 1046; DB 8; Length 3900;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TAGATCCTTCATAGAAAATGTGTTATTTCTCATCCAGACCAAGGGGCAACAGTTA 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
674 TAGATCCTTCATAGAAAATGTGTTATTTCTCATCCAGACCAAGGGGCAACAGTTA 733  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
61 ACAAACAAAATTTATGTTTTCATTTGAGATTAAAGGAAGTAAAGGAAGAAAAGATTAAAA 120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
734 ACAAACAAAATTTATGTTTTCATTTGAGATTAAAGGAAGTAAAGGAAGAAAAGATTAAAA 793  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
121 AAAATGCTTATCTCTTTGTTCTGTATATAATAATAAGAGACTTAAATCTTTAATAT 180  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
794 AAAATGCTTATCTCTTTGTTCTGTATATAATAATAAGAGACTTAAATCTTTAATAT 853  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
181 AATAATTTAGTTAGTTTCTAGTCATGAGCACCCTCAGAGACAGAGATTCAAGAAAA 240  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
854 AATAATTTAGTTAGTTTCTAGTCATGAGCACCCTCAGAGACAGAGATTCAAGAAAA 913  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
241 CAATTTTGTAAACATCTTATTAGAACTTTTGTAGTTAAAGTCTTGAAGTTAGAAATTAACA 300  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
914 CAATTTTGTAAACATCTTATTAGAACTTTTGTAGTTAAAGTCTTGAAGTTAGAAATTAACA 973  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
301 AAAAAATTACACAGAGAAACCAATACCCCTACCGTCAGGTTCATCATTAGGATGA 360  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
974 AAAAAAGTACACAGAGAAACCAATACCCCTACCGTCAGGTTCATCATTAGGATGA 1033  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
361 AATGTTTGTATATCAATTAATAACACACACAAAAATACATCTAATATATACAATATAT 420  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1034 AATGTTTGTATATCAATTAATAACACACACAAAAATACATCTAATATATACAATATAT 1093  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
421 GTTATACATATATTTTGTAAAACTTAGAGTTTTCAAAAACATCTAATATACATGATTAG 480  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1094 GTTATACATATATTTTGTAAAACTTAGAGTTTTCAAAAACATCTAATATACATGATTAG 1153  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
481 AGTTTATAGAAATACAAAATATTTAAAAAATATATATTTAAAAAACAATCTTAAGTCATT 540  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1154 AGTTTATAGAAATACAAAATATTTAAAAAATATATATTTAAAAAACAATCTTAAGTCATT 1213  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
541 CAGATCCTTCACACCTGTGTATCAATTTAGTCAATGTATGTATGATACAAATCAATTTAGTTC 600  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1214 CAGATCCTTCACACCTGTGTATCAATTTAGTCAATGTATGTATGATACAAATCAATTTAGTTC 1273  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
601 ACACAGAGTAAATTAATAGGATTAACCTAGGGAATATATATATATATATATATATATATAT 660  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1274 ACACAGAGTAAATTAATAGGATTAACCTAGGGAATATATATATATATATATATATATATAT 1333  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
661 AAAAAAGGGAATCAAAATTTAGAAATTTTGAATTTCCCAATGACACAACTCAACATGCAC 720  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1334 AAAAAAGGGAATCAAAATTTAGAAATTTTGAATTTCCCAATGACACAACTCAACATGCAC 1393  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
721 GCTGCCACCTCAGTCTCTCTCCACATGCTCTCAGTCTCTGCTTGGCTTTT 780  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1394 GCTGCCACCTCAGTCTCTCTCTCCACATGCTCTCAGTCTCTGCTTGGCTTTT 1453  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
781 TTCACATATGACAACTCGCCATGTCATGTGTGCCACGTGAGCTCTCTCTCTTCCCATGAT 840  
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||





ACCESSION AF193029  
 VERSION AF193029.1 GI:6684755  
 KEYWORDS Phaseolus vulgaris  
 SOURCE Phaseolus vulgaris  
 ORGANISM Phaseolus vulgaris  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 REFERENCE 1 (bases 1 to 2288)  
 AUTHORS Gerhardt, I.R. and Grossi de Sa, M.F.  
 TITLE Molecular characterization of a new arcelin-5 gene  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2288)  
 AUTHORS Gerhardt, I.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-OCT-1999) Embrapa/Cenargen, SAIN-Parque Rural-Final W5 Norte, Brasilia, DF 70770-900, Brazil  
 FEATURES  
 source Location/Qualifiers  
 1..2288 /organism="Phaseolus vulgaris"  
 /mol\_type="genomic DNA"  
 /cultivar="G02771"  
 /db\_xref="taxon:3885"  
 /clone="PIG53"  
 /tissue\_type="leaf"  
 <603..>1385  
 603..1385 /product="arcelin 5c"  
 /notes="arcs-III"  
 /codon\_start=1  
 /product="arcelin 5c"  
 /protein\_id="AAP23725.1"  
 /db\_xref="GI:6684756"  
 /translation="MASSNLLSLALFLVLLTHANSATETSPNFTSFHPDDPKMLQSD ATISTKQLLITSYELSRVDSLGRALYSDFIQIKDNNVASFDTKFTFIIRPETNGN SAYGLALVPGSKPGKPYLGIFNDITPEPDARTVAVFNLNRIDIDVNAIKP YANEKCPFKYNGSKTVOITYDSSKNDLRLVFLHTFVSQVKCSVSATVQLEKEVNEWV SVGFSATSGLTZENTETHDVLWSFSKPRKLSNLLNLL"  
 mRNA  
 CDS  
 Query Match 6.0%; Score 69; DB 8; Length 2288;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-24;  
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 965 TCTCCATAAATATCTATTAAATTTAACTAACTAATTTATATATATCTTTTGTAGAGGTG 1024  
 Db TCTCCATAAATATCTATTAAATTTAACTAACTAATTTATATATATCTTTTGTAGAGGTG 1024  
 Qy 1025 GATGCATTG 1033  
 Db 459 GATGCATTG 467  
 RESULT 10  
 PVPDLEC2  
 LOCUS PVPDLEC2 1992 bp DNA linear PLN 21-MAR-1995  
 DEFINITION P. vulgaris cv. Pinto Palec2 gene for phytohemagglutinin (PHA).  
 ACCESSION X04659  
 VERSION X04659.1 GI:21022  
 KEYWORDS glycoprotein; Pdlec2 gene; phytohemagglutinin; phytohemagglutinin-L; signal peptide.  
 SOURCE Phaseolus vulgaris  
 ORGANISM Phaseolus vulgaris  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 REFERENCE 1 (bases 1 to 1992)  
 AUTHORS Voelker, T.A., Staswick, P. and Chrispeels, M.J.  
 TITLE Molecular analysis of two phytohemagglutinin genes and their expression in Phaseolus vulgaris cv. Pinto, a lectin-deficient cultivar of the bean

JOURNAL EMBO J. 5, 3075-3082 (1986)  
 COMMENT See also <X04660> for dlec1 gene (pseudogene). Several inverted repeats are described in the 5' upstream region of the PHA coding region.  
 2 pot. glycosylation sites are found at the AA residues 33 and 92. Pdlec2 mRNA is reduced 20-fold (with respect to mRNA levels in normal cultivars); a 100bp deletion between pos. 501 and 502 (compared to dlec1, dlec2 and Palec1) containing a large tandem repeat may be responsible for the low level of expression of Pdlec2.  
 FEATURES  
 source Location/Qualifiers  
 1..1992 /organism="Phaseolus vulgaris"  
 /mol\_type="genomic DNA"  
 /strain="Pinto Uilll"  
 /db\_xref="taxon:3885"  
 /tissue\_type="young leaves (greenhouse)."  
 742..747  
 promoter /note="put. TATA-box"  
 785..1606  
 CDS /note="unnamed protein product; precursor polypeptide (AA -21 to 252)"  
 /codon\_start=1  
 /protein\_id="CAA28362.1"  
 /db\_xref="GI:21023"  
 /db\_xref="GOA:P15231"  
 /db\_xref="SWISS-PROT:P15231"  
 /translations="MASSNLLSLALFLVLLTHANSASQTFFSDRFNETNLLIQGDAS VSSGQLRUTNVNSGEPTVSGLRAPYQIWDYITTYGNVASFDTNFTFNILVNN AGPADGLAFALVPGSKPGKPYLGIFNDITPEPDARTVAVFNLNRIDIDVNAIKP IDVNSIKSITKTPDFVNGENAEVHTYESTKLLVASLYPSLKTSTFTVSDIVDLKS VLPEWISGVGSATTTGTTKGNVETNLSWFSKLSGDTTSEGINLANLVINQIL"  
 sig\_peptide 785..847  
 /note="put. signal peptide (AA -21 to -1)"  
 mat\_peptide 848..1603  
 /product="mature PHA (AA 1-252)"  
 misc\_feature 1691..1699  
 /note="put. polyA signal"  
 ORIGIN  
 Query Match 5.7%; Score 65; DB 8; Length 1992;  
 Best Local Similarity 100.0%; Pred. No. 8.9e-22;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1084 AAAAAAGTTGGAAGATTTTGCATTTTGTGTTTGTATAAATAGAGAAGAGTGTATGTTA 1143  
 Db 709 AAAAAAGTTGGAAGATTTTGCATTTTGTGTTTGTATAAATAGAGAAGAGTGTATGTTA 768  
 Qy 1144 ATGCA 1148  
 Db 769 ATGCA 773  
 RESULT 11  
 PHVLECT  
 LOCUS PHVLECT 1689 bp DNA linear PLN 27-APR-1993  
 DEFINITION P. vulgaris lectin gene, complete cds.  
 ACCESSION J01261  
 VERSION J01261.1 GI:169354  
 KEYWORDS lectin.  
 SOURCE Phaseolus vulgaris  
 ORGANISM Phaseolus vulgaris  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 REFERENCE 1 (bases 257 to 1115)  
 AUTHORS Hoffman, I.M., Ma, Y. and Barker, R.F.  
 TITLE Molecular cloning of Phaseolus vulgaris lectin mRNA and use of cDNA as a probe to estimate lectin transcript levels in various tissues  
 JOURNAL Nucleic Acids Res. 10 (23), 7819-7828 (1982)  
 MEDLINE 83116994  
 PUBMED 6897567

```

REFERENCE
AUTHORS      Hoffman,L.M.
TITLE        Structure of a chromosomal Phaseolus vulgaris lectin gene and its
JOURNAL      J. Mol. Appl. Genet. 2 (5), 447-453 (1984)
MEDLINE      85008540
PUBMED       6090563
COMMENT      Original source text: P.vulgaris cv. Tendergreen DNA, clone
              p-lambda-lec5.7 [2]; and cotyledon, cDNA to mRNA, clone pPVL134
              [1].
CAT and TATA boxes are located at positions 165-169 and 226-231
respectively. Two tandem poyadenylation signals are present at
position 1084-1093.
A magnetic tape of this sequence and a draft entry were kindly
provided by L.M.Hoffman (23-OCT-1985).

FEATURES
source       1..1689
             /organism="Phaseolus vulgaris"
             /mol_type="genomic DNA"
             /db_xref="taxon:3885"
mRNA         255..1115
             /product="lectin mRNA"
CDS          265..1005
             /note="lectin prepeptide"
             /codon_start=1
             /protein_id="AAA33769.1"
             /db_xref="GI:169355"
             /translation="MIMASSKLSLALFLALLSHANSATSFIIIDAFNKNLILQGD
              ATVSNGLQISVNSYDMSRAFYSAPIQIRDSITGNVASTPTNFTNIRHRCANSA
              VGLDFVLVPQESKGDVTVEPTFLSRISIDVNNNDIKSPKVDYDQONAEVFI
              TYNSTKRVFSLVSNPSTGKSNVSTTVELEKEVDVWVGFSATSGAYQWSYTHDV
              LSWFSFKFINKKQKRSNIVLNKIL"
sig_peptide  265..333
             /note="lectin signal peptide"
mat_peptide  334..1002
             /product="lectin mature peptide"
ORIGIN       71 bp upstream AhalII site.

Query Match      4.3%; Score 49; DB 8; Length 1689;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1100 TTTTCATTGTGTGTATTAATAGAGAGAGAGTGTGTTAATGCA 1148
      |||||||||||||||||||||||||||||||||||||||||||
DB 209 TTTTCATTGTGTGTATTAATAGAGAGAGTGTGTTAATGCA 257

RESULT 12
PHVARC1A      PHVARC1A      4564 bp DNA linear PLN 27-APR-1993
LOCUS         Phaseolus vulgaris arcelin (arc) gene, complete cds.
DEFINITION    M68913
ACCESSION     M68913.1 GI:169313
VERSION       arcelin; lectin-like seed protein.
KEYWORDS      Phaseolus vulgaris
SOURCE        Phaseolus vulgaris
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
              Phaseolus.
REFERENCE     1 (bases 1 to 4564)
AUTHORS       Anthony,J.L., Vonder Haar,R.A. and Hall,T.C.
TITLE         Nucleotide sequence of a genomic clone encoding arcelin, a
              lectin-like seed protein from Phaseolus
JOURNAL       Plant Physiol. 97, 839-840 (1991)
COMMENT      Original source text: Phaseolus vulgaris (library: Lambda zap II)
              young plant leaf DNA.
FEATURES
source       1..4564
             /organism="Phaseolus vulgaris"
             /mol_type="genomic DNA"
             /db_xref="taxon:3885"
             /tissue_type="leaf"

/dev_stage="young plant"
/tissue_lib="Lambda Zap II"
3332..4172
/genes="arc"
3332..3338
/genes="arc"
3361..3364
/genes="arc"
/note="cap region"
3375..4172
/genes="arc"
/codon_start=1
/product="arcelin"
/protein_id="AAA33753.1"
/db_xref="GI:169314"
/translation="MASSNLTLLALFLVLLTHANSNDASFNVEFNKTNLILQGDAT
VSSECHLLTNVKGNEEDSGRAFYSAPIQIRDTIDNLSFSTNFTNIRHRCANSA
AYGLFALVPGSRPKLGRILGFTNTNDRHVAVVDVTSNRIEDVNSIRPI
ATSCNFGNNGEKARVITYDSPKNDLRVSLYPSSSEKCHVSATVPLEKEVEDWVS
VGFSATSGKKKETTETHNVLWSFSSNFNFKGKRSNIVLNKIL"
3375..3437
/genes="arc"
3438..4169
/genes="arc"
/product="arcelin"

ORIGIN
Query Match      4.2%; Score 48; DB 8; Length 4564;
Best Local Similarity 100.0%; Pred. No. 4.6e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1068 AATAATAAATGAAGAAAAAGTTGGAAGATTTGCAATTTGTTGTT 1115
      |||||||||||||||||||||||||||||||||||||||||||
DB 3283 AATAATAAATGAAGAAAAAGTTGGAAGATTTGCAATTTGTTGTT 3330

RESULT 13
PVPDLEC1      PVPDLEC1      1768 bp DNA linear PLN 12-SEP-1993
LOCUS         P. vulgaris cv. Pinto pseudogene Pdlec1 for phytohemagglutinin
DEFINITION    (PHA).
ACCESSION     X04660
VERSION       X04660.1 GI:21020
KEYWORDS      Pdlec1 gene; phytohemagglutinin; pseudogene.
SOURCE        Phaseolus vulgaris
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
              Phaseolus.
REFERENCE     1 (bases 1 to 1768)
AUTHORS       Voelker,T.A., Staswick,P. and Chrispeels,M.J.
TITLE         Molecular analysis of two phytohemagglutinin genes and their
              expression in Phaseolus vulgaris cv. Pinto, a lectin-deficient
              cultivar of the bean
JOURNAL       EMBO J. 5, 3075-3082 (1986)
COMMENT      See also <X04659> for Pdlec2 gene.
              Several inverted repeats are described in the 5' upstream region of
              the PHA gene.
              None of the 3 possible reading frames allow the translation of its
              message into a complete PHA polypeptide. The PHA translational
              start codon (pos. 708-710) starts a RF which codes for a truncated
              50AA polypeptide with a N-terminal sequence completely identical to
              the dlec1 gene product, but is out of register after 10 codons due
              to a single bp deletion (pos. 739/740).
FEATURES
source       1..1768
             /organism="Phaseolus vulgaris"
             /mol_type="genomic DNA"
             /strain="Pinto U111"
             /db_xref="taxon:3885"
             /tissue_type="young leaves (greenhouse)."
             665..670
             promoter

```

CDS

/note="put. TATA-box"  
708..866  
/note="unnamed protein product; pseudogene region"  
/codon\_start=1  
/protein\_id="CAA28363.1"  
/db\_xref="GI:21021"  
/db\_xref="REMBT:CAA28363"  
/translation="MASSNLSLALSLCFSPQTQPAKPSSASIGSTKPTLSSKAMPP  
SHPPASVD"  
739..740  
variation  
/note="tt is tct in pDlec2 (c deletion, causing a frame  
shift mutation)"  
1620..1635  
misc\_feature  
/note="put. polyA signal"

ORIGIN

Query Match 4.0%; Score 46; DB 8; Length 1768;  
Best Local Similarity 100.0%; Pred. No. 5.3e-12;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1103 TGCATTGTTGTTGTTATAATAGAGAGAGAGTGTGTTAATGCA 1148  
|||  
DB 651 TGCATTGTTGTTGTTATAATAGAGAGAGAGTGTGTTAATGCA 696  
|||

RESULT 14

AF255723 1852 bp DNA linear PLN 24-MAY-2000  
LOCUS  
DEFINITION Phaseolus vulgaris ARL5-IV pseudogene, complete sequence.  
ACCESSION AF255723  
VERSION AF255723.1 GI:8050712  
KEYWORDS  
SOURCE Phaseolus vulgaris  
ORGANISM Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;  
Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.

REFERENCE 1 (bases 1 to 1852)  
AUTHORS Gerhardt, I.R., Pappas, G., Chrispeels, M.J. and Grossi de Sa, M.F.  
TITLE Direct Submission  
JOURNAL Submitted (13-APR-2000) Biotechnology, Embrapa-Cenargen, SAIN  
Parque Rural- Final W5 Norte, CP 02372, Brasilia, DF 70 770-900,  
Brazil

FEATURES  
source Location/Qualifiers  
1..1852  
/organism="Phaseolus vulgaris"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3885"  
146..814  
/genes="ARL5-IV"  
/pseudo

gene

ORIGIN

Query Match 3.7%; Score 42; DB 8; Length 1852;  
Best Local Similarity 100.0%; Pred. No. 6.1e-10;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1098 GATTTCGATTTCTGTTGTTATAATAGAGAGAGTGTG 1139  
|||  
DB 107 GATTTCGATTTCTGTTGTTATAATAGAGAGAGTGTG 148  
|||

RESULT 15

E09281 1211 bp DNA linear PAT 29-SEP-1997  
LOCUS  
DEFINITION DNA encoding lectin-like protein of kidney bean.  
ACCESSION E09281  
VERSION E09281.1 GI:22025907  
KEYWORDS JP 1995132092-A/1.  
SOURCE unidentified  
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 1211)  
AUTHORS Hagiwara, K.  
TITLE NEW KIDNEY BEAN GENE  
JOURNAL Patent: JP 1995132092-A 1 23-MAY-1995;  
NORIN SUIJANSYO NOGYO SEIBUTSU SHIGEN KENKYUSHO  
COMMENT OS Phaseolus sp. (kidney bean)  
PN JP 1995132092-A/1  
PD 23-MAY-1995  
PF 11-NOV-1993 JP 1993305988  
PI HAGIWARA KIYOSHI  
PC CL2N15/09, CL2N1/21//C07K14/42, (CL2N1/21, CL2R1/19); CC  
strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FH Key Location/Qualifiers  
FH 1..1211  
FT source /organism="Phaseolus sp."  
FT /strain="Kentucky wonder".  
FT Location/Qualifiers  
1..1211  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

FEATURES  
source Location/Qualifiers  
1..1211  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

ORIGIN

Query Match 3.6%; Score 41; DB 6; Length 1211;  
Best Local Similarity 100.0%; Pred. No. 2.1e-09;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1108 TTGTTGTTGTTATAATAGAGAGAGTGTGTTAATGCA 1148  
|||  
DB 263 TTGTTGTTGTTATAATAGAGAGAGTGTGTTAATGCA 303  
|||

Search completed: June 2, 2004, 06:17:34  
Job time : 4660 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 03:34:13 ; Search time 108 Seconds  
(without alignments)  
5898.922 Million cell updates/sec

Title: US-10-015-637-1  
Perfect score: 1148  
Sequence: 1 taggacccctcaataagaaaa.....agagagtgtgtaataatgca 1148

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS-COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.8	7430	4	US-08-976-259-64 Sequence 64, Appl
2	20	1.7	802	4	US-09-807-258-19 Sequence 19, Appl
3	20	1.7	1597	4	US-09-071-035-87 Sequence 87, Appl
4	20	1.7	1677	4	US-09-134-000C-2844 Sequence 2844, Ap
5	20	1.7	1680	4	US-09-071-035-85 Sequence 85, Appl
6	20	1.7	3627	1	US-08-104-072B-6 Sequence 6, Appl
7	20	1.7	3627	1	US-08-351-413-7 Sequence 7, Appl
8	20	1.7	3627	2	US-09-025-583-7 Sequence 7, Appl
9	20	1.7	392000	4	US-10-027-983-11 Sequence 11, Appl
10	20	1.7	640681	4	US-09-790-988-1 Sequence 1, Appl
11	20	1.7	1230025	4	US-09-198-452A-1 Sequence 1, Appl
12	19	1.7	181	4	US-03-669-751-213 Sequence 213, App
13	19	1.7	1212	4	US-03-489-039A-2871 Sequence 2871, Ap
14	19	1.7	1776	4	US-09-134-001C-391 Sequence 391, App
15	19	1.7	2590	4	US-09-023-655-301 Sequence 301, App
16	19	1.7	4105	4	US-09-634-238-182 Sequence 182, App
17	19	1.7	4784	4	US-09-634-238-24 Sequence 24, Appl
18	19	1.7	30246	4	US-08-956-171E-56 Sequence 56, Appl
19	19	1.7	38653	4	US-09-922-445-1 Sequence 1, Appl
20	19	1.7	319608	4	US-09-539-333D-1 Sequence 1, Appl
21	19	1.7	319608	4	US-09-679-409-1 Sequence 1, Appl
22	19	1.7	640681	4	US-09-790-988-1 Sequence 1, Appl
23	18	1.6	109	4	US-09-539-601-45 Sequence 45, Appl
24	18	1.6	115	4	US-09-539-601-50 Sequence 50, Appl
25	18	1.6	115	4	US-09-539-601-51 Sequence 51, Appl
26	18	1.6	300	4	US-08-956-171E-3111 Sequence 3111, Ap
27	18	1.6	311	4	US-09-539-401C-59 Sequence 59, Appl

Sequence 157, App  
Sequence 157, App  
Sequence 157, App  
Sequence 157, App  
Sequence 157, App  
Sequence 157, App  
Sequence 176, App  
Sequence 176, App  
Sequence 15, Appl  
Sequence 937, App  
Sequence 1926, Ap  
Sequence 3, Appl  
Sequence 29, Appl  
Sequence 160, App  
Sequence 161, App  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 90, Appl  
Sequence 5, Appl

US-08-976-259-64  
; Sequence 64, Application US/08976259  
; Patent No. 6316609  
; GENERAL INFORMATION:  
; APPLICANT: Dillion, Patrick J.  
; APPLICANT: Choi, Gil H.  
; APPLICANT: Welch, Rodney A.  
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli  
; Patent No. 6316609  
; NUMBER OF SEQUENCES: 142  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Ave, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,259  
; FILING DATE: Herewith  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7430 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-976-259-64

ALIGNMENTS

RESULT 1  
US-08-976-259-64  
; Sequence 64, Application US/08976259  
; Patent No. 6316609  
; GENERAL INFORMATION:  
; APPLICANT: Dillion, Patrick J.  
; APPLICANT: Choi, Gil H.  
; APPLICANT: Welch, Rodney A.  
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli  
; Patent No. 6316609  
; NUMBER OF SEQUENCES: 142  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Ave, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,259  
; FILING DATE: Herewith  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7430 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-976-259-64

Query Match 1.8%; Score 21; DB 4; Length 7430;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 ATTTAAATAATATTTTAA 520  
|||||

Db 7289 ATTAAATAATATAATTTAA 7309

RESULT 2

US-09-807-258-19

Sequence 19, Application US/09807258

Patent No. 6670166

GENERAL INFORMATION:

APPLICANT: E. I. du Pont de Nemours and Company

TITLE OF INVENTION: Arthropod Protein Disulfide Isomerases

FILE REFERENCE: BB-1253 PCT

CURRENT APPLICATION NUMBER: US/09/807,258

CURRENT FILING DATE: 2001-06-11

PRIOR APPLICATION NUMBER: 60/104,376

PRIOR FILING DATE: 1998-10-15

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Microsoft Office 97

SEQ ID NO 19

LENGTH: 802

TYPE: DNA

ORGANISM: Hottentotta judiaca

US-09-807-258-19

Query Match 1.7%; Score 20; DB 4; Length 802;

Best Local Similarity 100.0%; Pred. No. 8.3; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1070 TAATAAAATCAAGAAAGAAA 1089

Db 771 TAATAAAATCAAGAAAGAAA 790

RESULT 3

US-09-071-035-87

Sequence 87, Application US/09071035

Patent No. 6443043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 1597 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-87

Query Match 1.7%; Score 20; DB 4; Length 1597;

Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 AAAATCAAAATTAGAATTTT 689

Db 1112 AAAATCAAAATTAGAATTTT 1131

RESULT 4

US-09-134-000C-2844

Sequence 2844, Application US/09134000C

Patent No. 6617156

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-032

CURRENT APPLICATION NUMBER: US/09/134,000C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/055,778

PRIOR FILING DATE: 1997-08-15

NUMBER OF SEQ ID NOS: 6812

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2844

LENGTH: 1677

TYPE: DNA

ORGANISM: Enterococcus faecalis

US-09-134-000C-2844

Query Match 1.7%; Score 20; DB 4; Length 1677;

Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 AAAATCAAAATTAGAATTTT 689

Db 1189 AAAATCAAAATTAGAATTTT 1208

RESULT 5

US-09-071-035-85

Sequence 85, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 1597 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-87

```
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-071-035-85

Query Match          1.7%; Score 20; DB 4; Length 1680;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 670 AAAATCAATTAGAATTTT 639
Db 1192 AAAATCAATTAGAATTTT 1211

RESULT 6
US-08-104-072B-6/c
; Sequence 6, Application US/08104072B
; Patent No. 5639948
; GENERAL INFORMATION:
; APPLICANT: Michiels, Frank
; APPLICANT: Morioka, Sinji
; APPLICANT: Scheirlinck, Trees
; APPLICANT: Komari, Toshihiko
; TITLE OF INVENTION: Stamen-specific Promoters from Rice
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5639948west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/104,072B
; FILING DATE: 05-AUG-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 9200272
; FILING DATE: 06-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91403352.7
; FILING DATE: 10-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91402590.3
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91400318.1
; FILING DATE: 08-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.93USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Oryza sativa

; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..2845
; OTHER INFORMATION: /function= "sequence comprising
; OTHER INFORMATION: anther-specific PT72 promoter"
; FEATURE:
; NAME/KEY: TATA signal
; LOCATION: 2733..2739
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2765
; OTHER INFORMATION: /product= "transcription
; OTHER INFORMATION: initiation"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2846
; OTHER INFORMATION: /product= "ATG start translation of
; OTHER INFORMATION: T72 gene"
US-08-104-072B-6

Query Match          1.7%; Score 20; DB 1; Length 3627;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1071 AATAAATGAAGAAAAAG 1090
Db 991 AATAAATGAAGAAAAAG 972

RESULT 7
US-08-351-413-7/c
; Sequence 7, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2045
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Oryza sativa  
STRAIN: Akihikari  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..2845  
OTHER INFORMATION: /label= PT72  
OTHER INFORMATION: /note= "sequence comprising anther specific  
OTHER INFORMATION: promoter PT72"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2733..2739  
OTHER INFORMATION: /label= TATA  
OTHER INFORMATION: /note= "TATA Box"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2765  
OTHER INFORMATION: /note= "transcription initiation  
OTHER INFORMATION: determined by primer extension"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2846..2848  
OTHER INFORMATION: /label= ATG  
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"  
US-08-351-413-7

Query Match 1.7%; Score 20; DB 1; Length 3627;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1071 AATAAATGAAGAAAAAG 1090  
|||||  
Db 991 AATAAATGAAGAAAAAG 972

RESULT 8  
US-09-025-583-7/c  
Sequence 7, Application US/09025583  
Patent No. 5977433  
GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
APPLICANT: Leemans, Jan  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/025.583  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/351.413  
FILING DATE:  
APPLICATION NUMBER: US 07/899,072  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/970,849  
FILING DATE: 03-NOV-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-102PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3627 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Oryza sativa  
STRAIN: Akihikari  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..2845  
OTHER INFORMATION: /label= PT72  
OTHER INFORMATION: /note= "sequence comprising anther specific  
OTHER INFORMATION: promoter PT72"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2733..2739  
OTHER INFORMATION: /label= TATA  
OTHER INFORMATION: /note= "TATA Box"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2765  
OTHER INFORMATION: /note= "transcription initiation  
OTHER INFORMATION: determined by primer extension"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2846..2848  
OTHER INFORMATION: /label= ATG  
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"  
US-09-025-583-7

Query Match 1.7%; Score 20; DB 2; Length 3627;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1071 AATAAATGAAGAAAAAG 1090  
|||||  
Db 991 AATAAATGAAGAAAAAG 972

RESULT 9  
US-10-027-983-11/c  
Sequence 11, Application US/10027983  
Patent No. 6617162  
GENERAL INFORMATION:  
APPLICANT: Kenneth W. Dobie  
APPLICANT: Mark P. Roach  
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION  
FILE REFERENCE: RTS-0340  
CURRENT APPLICATION NUMBER: US/10/027,983  
CURRENT FILING DATE: 2001-12-18  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 11  
LENGTH: 392000  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 137740  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 137742

```
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (138122)...(138221)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 145507
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 151967
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (151967)...(1542066)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 154217
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (164037)...(164136)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (174657)...(174756)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (186224)...(186323)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (195242)...(195341)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 202703
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (202771)...(202870)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (206246)...(215602)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (218126)...(218225)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (220360)...(220459)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (222717)...(222816)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (223581)...(224080)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (227487)...(227586)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (230157)...(230256)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (232299)...(232398)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (236552)...(236651)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (238789)...(248788)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: exon
LOCATION: (118288)...(119101)
OTHER INFORMATION: exon 1C
NAME/KEY: exon:intron junction
LOCATION: (151129)...(151130)
OTHER INFORMATION: exon 5:intron 5
NAME/KEY: exon:intron junction
LOCATION: (299248)...(299249)
OTHER INFORMATION: exon 9:intron 9
```

```
NAME/KEY: exon:intron junction
LOCATION: (348578)...(348579)
OTHER INFORMATION: exon 10:intron 10
NAME/KEY: intron
LOCATION: (348579)...(381838)
OTHER INFORMATION: intron 10
NAME/KEY: intron:exon junction
LOCATION: (386185)...(386186)
OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11
```

```
Query Match 1.7%; Score 20; DB 4; Length 392000;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1071 AATAAATGAAAGAAAAAG 1090
|||||
DB 276309 AATAAATGAAAGAAAAAG 276290
```

```
RESULT 10
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1
```

```
Query Match 1.7%; Score 20; DB 4; Length 640681;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 499 TATTTAAAAATATAATTTT 518
|||||
DB 186362 TATTTAAAAATATAATTTT 186381
```

```
RESULT 11
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(15000)
; OTHER INFORMATION: n=a or c or g or t
```



[illegible]

OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (750001)..(765000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (765001)..(780000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (780001)..(795000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (795001)..(810000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (810001)..(825000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (825001)..(840000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (840001)..(855000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (855001)..(870000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (870001)..(885000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (885001)..(900000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (900001)..(915000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature

Query Match 1.7%; Score 20; DB 4; Length 1230025;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 GAAGAAAAAGATTAAAAA 122  
Db 1175674 GAAGAAAAAGATTAAAAA 1175655

RESULT 12  
US-09-669-751-213  
Sequence 213, Application US/09569751  
Patent No. 6551575  
GENERAL INFORMATION:  
APPLICANT: Greenspan, Ralph J.  
TITLE OF INVENTION: Methods for Identifying Compounds for Motion Sickness, Vertigo and Other Disorders Related to Balance and the Perception of Gravity  
FILE REFERENCE: P-NI 3864  
CURRENT APPLICATION NUMBER: US/09/669,751  
CURRENT FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US 60/168,579  
PRIOR FILING DATE: 1999-12-02  
NUMBER OF SEQ ID NOS: 261  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 213  
LENGTH: 181  
TYPE: DNA  
ORGANISM: Drosophila  
US-09-669-751-213

Query Match 1.7%; Score 19; DB 4; Length 181;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 AAATATAACACACAAAA 396  
|||||

Db 107 AAATATAACACACAAAA 125

RESULT 13  
US-09-489-039A-2871/c  
Sequence 2871, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 2871  
LENGTH: 1212  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2871

Query Match 1.7%; Score 19; DB 4; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 ACAATTAAATAAAAAAGGG 669  
|||||  
Db 47 ACAATTAAATAAAAAAGGG 29

RESULT 14  
US-09-134-001C-391/c  
Sequence 391, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 391  
LENGTH: 1776  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-391

Query Match 1.7%; Score 19; DB 4; Length 1776;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 TTAATATATATATTTGTAAT 192  
|||||  
Db 547 TTAATATATATATTTGTAAT 529

RESULT 15  
US-09-023-655-301  
Sequence 301, Application US/09023655  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE EXPRESSION

```

;
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCITE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/05/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 301:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2590 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LATRUT02
; CLONE: 1353065
; US-09-023-655-301

```

```

Query Match 1.7%; Score 19; DB 4; Length 2590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1068 AATAATAAATGAAAGAAA 1086
DB 2265 AATAATAAATGAAAGAAA 2283

```

```

Search completed: June 2, 2004, 07:16:20
Job time : 117 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:36:30 ; Search time 3393 Seconds  
(without alignments)  
10133.545 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148  
Sequence: 1 taggtcttcctcaagaaaa.....agagagtgtggttaatsca 1148

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	25	2.2	846	28	BH570165 BOGPE54TR
c 2	24	2.1	194	29	CE710533 tigr-gss-
c 3	24	2.1	491	29	BX122318 Danio rer
c 4	24	2.1	536	13	BX734218 BX734218

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	24	2.1	696	28	BZ882476
6	24	2.1	733	13	BX693278
c 7	24	2.1	820	29	CC474046
c 8	24	2.1	1509	29	CG755596
c 9	23	2.0	259	14	R95629
c 10	23	2.0	307	9	AA380112
c 11	23	2.0	449	14	CB378526
c 12	23	2.0	543	29	TA191802P
c 13	23	2.0	604	29	CE607569
c 14	23	2.0	611	19	CE223942
c 15	23	2.0	654	14	CB935043
c 16	23	2.0	747	29	CC586297
c 17	23	2.0	786	28	BZ163794
c 18	22	1.9	53	9	A1561282
c 19	22	1.9	162	14	TL4440
c 20	22	1.9	168	9	AA528923
c 21	22	1.9	171	9	AA525589
c 22	22	1.9	199	14	CD072757
c 23	22	1.9	201	9	AW017393
c 24	22	1.9	210	9	AA801591
c 25	22	1.9	210	9	A1977469
c 26	22	1.9	218	9	AA508323
c 27	22	1.9	221	9	AA801590
c 28	22	1.9	225	28	BH439146
c 29	22	1.9	240	14	R95407
c 30	22	1.9	245	9	AA528868
c 31	22	1.9	246	9	A1977349
c 32	22	1.9	250	14	R95403
c 33	22	1.9	300	14	R95405
c 34	22	1.9	306	9	A1976001
c 35	22	1.9	308	9	A1977237
c 36	22	1.9	310	14	R95404
c 37	22	1.9	326	14	CD171263
c 38	22	1.9	345	14	R95607
c 39	22	1.9	350	9	A1976293
c 40	22	1.9	359	14	CA905893
c 41	22	1.9	360	14	CD190460
c 42	22	1.9	368	9	AW017025
c 43	22	1.9	370	9	A1067007
c 44	22	1.9	372	9	A1976918
c 45	22	1.9	372	9	AW017201

## ALIGNMENTS

RESULT 1  
BH570165/c  
LOCUS  
DEFINITION  
BOGPE54TR BOGP Brassica oleracea genomic clone BOGPE54, genomic survey sequence.  
ACCESSION  
BH570165  
VERSION  
BH570165.1 GI:17822019  
KEYWORDS  
GSS.  
SOURCE  
Brassica oleracea  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE  
1 (bases 1 to 846)  
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
COMMENT  
Other\_GSSs: BOGPE54TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

BH570165 846 bp DNA linear GSS 14-DEC-2001  
BOGPE54TR BOGP Brassica oleracea genomic clone BOGPE54, genomic survey sequence.

BH570165  
BH570165.1 GI:17822019  
GSS.  
Brassica oleracea  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 846)  
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
COMMENT  
Other\_GSSs: BOGPE54TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

```

FEATURES
  source
    Location/Qualifiers
      1..846
      /organism="Brassica oleracea"
      /mol_type="genomic DNA"
      /strain="T01000DH3"
      /db_xref="taxon:3712"
      /clone="BOGPE54"
      /clone_lib="BOGP"
      /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
  Query Match      2.2%; Score 25; DB 28; Length 846;
  Best Local Similarity 100.0%; Pred. No. 3.4;
  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 ATTAAAAAATATAATTTTAAAAA 524
    |||||
Db 832 ATTAAAAAATATAATTTTAAAAA 808

RESULT 2
CE710533/c
LOCUS
DEFINITION
  tigr-gss-dog-17000369461433 Dog Library Canis familiaris genomic,
  genomic survey sequence.
ACCESSION
  CE710533
VERSION
  CE710533.1 GI:37029944
KEYWORDS
  GSS:
SOURCE
  Canis familiaris (dog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
  1 (bases 1 to 194)
  Kirknes,E.P., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
  Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
  Venter,J.C.
  The dog genome: survey sequencing and comparative analysis
  Science 301 (5641), 1898-1903 (2003)
  MEDLINE
  22875432
  PUBMED
  14512627
  Contact: Kirknes EP
  The Institute for Genomic Research
  Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
  Rockville, MD 20850, USA
  Tel: 301-838-0200
  Fax: 301-838-0208
  Email: ekirknes@tigr.org
  Class: shotgun.
FEATURES
  source
    Location/Qualifiers
      1..194
      /organism="Canis familiaris"
      /mol_type="genomic DNA"
      /strain="Standard Poodle"
      /db_xref="taxon:9615"
      /clone_lib="DOG Library"
      /note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN
  Query Match      2.1%; Score 24; DB 29; Length 194;
  Best Local Similarity 100.0%; Pred. No. 13;
  Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 TATAACAATATATGTTATACAT 431
    |||||
Db 155 TATAACAATATATGTTATACAT 132

RESULT 3
BX122318
LOCUS
DEFINITION
  Danio rerio genomic clone DKEY-72P11, genomic survey sequence.

FEATURES
  source
    Location/Qualifiers
      1..491
      /organism="Danio rerio"
      /mol_type="genomic DNA"
      /db_xref="taxon:7955"
      /clone="DKEY-72P11"
      /tissue_type="Testis"
      /note="vector pindigoBAC-536"

ORIGIN
  Query Match      2.1%; Score 24; DB 29; Length 491;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 CAATATTTAAAAAATAATTTT 518
    |||||
Db 89 CAATATTTAAAAAATAATTTT 112

RESULT 4
BX734218
LOCUS
DEFINITION
  BX734218 XGC-tadpole Silurana tropicalis cDNA clone TTPA076k17 5',
  mRNA sequence.
ACCESSION
  BX734218
VERSION
  BX734218.1 GI:38406959
KEYWORDS
  EST.
SOURCE
  Silurana tropicalis (western clawed frog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
  Xenopodinae; Silurana.
  1 (bases 1 to 536)
  Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
  Sanger Xenopus tropicalis EST project 2001 (11_2003)
  Unpublished (2003)
  Contact: Croning MDR
  Sanger Institute
  Hinxton, Cambridgeshire, CB10 1SA, UK
  Email: trop@sanger.ac.uk
  Sanger Xenopus tropicalis EST project 2001
  TROPICALIS_SEQUENCE_ID: TTPA076k17.plkSP6
  Sequencing primer: SP6
  This sequence is from a Xenopus Gene Collection (XGC) library
  constructed by Nigel Garrett.
  cDNA was oligo dt primed from 5ug of poly A+ RNA from tadpole
  embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
  EcoRI at the 5' end and NotI at the 3' end.
  Vector: pCS107; Site1: EcoRI; Site2: NotI
  Host: Escherichia coli DH10B.
FEATURES
  source
    Location/Qualifiers
      1..536
      /organism="Silurana tropicalis"
      /mol_type="mRNA"
      /db_xref="taxon:8364"

```

```

/clone="TtpA076k17"
/dev stage="tadpole (stage 35-40)"
/lab_host="E. coli DH10B"
/clone_lib="XGC-tadpole"
/notes="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN
Query Match      2.1%; Score 24; DB 13; Length 536;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 630 TAGGGAATATATATATATATACA 653
Db 160 TAGGGAATATATATATATATACA 183

RESULT 5
BZ882476
LOCUS
DEFINITION
CH240_254K15_TV CHORI-240 Bos taurus genomic clone CH240_254K15,
genomic survey sequence.
ACCESSION
BZ882476
VERSION
BZ882476.1 GI:29109878
SOURCE
Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 696)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Chen,D., Riggs,F., de Jong,P.,
Crawford,A.M. and McGowan,J.C.
Bovine BAC End Sequences from Library CHORI-240
Unpublished (2003)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.html). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.choi.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by AgResearch Ltd., New Zealand and The
Institute of Genomic Research (TIGR), USA.
Plate: 254 row: K column: 15
Seg primer: T7
Class: BAC ends.

FEATURES
Location/Qualifiers
source
1..696
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_254K15"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/notes="vector: pIARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match      2.1%; Score 24; DB 28; Length 696;
Best Local Similarity 100.0%; Pred. No. 9.9;

```

```

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 ATAACACACACAAAATACATCTA 405
Db 351 ATAACACACACAAAATACATCTA 374

RESULT 6
BX693278/c
LOCUS
DEFINITION
BX693278 XGC-neurula Silurana tropicalis cDNA clone TNeu03f10 3',
mRNA sequence.
ACCESSION
BX693278
VERSION
BX693278.1 GI:38342398
SOURCE
EST.
Silurana tropicalis (western clawed frog)
ORGANISM
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
REFERENCE
1 (bases 1 to 733)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu03f10.q1kT7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.

FEATURES
Location/Qualifiers
source
1..733
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu03f10"
/dev stage="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match      2.1%; Score 24; DB 13; Length 733;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 630 TAGGGAATATATATATATACA 653
Db 448 TAGGGAATATATATATATACA 425

RESULT 7
CC474046
LOCUS
DEFINITION
CH240_298D10_TARBAC13P2 CHORI-240 Bos taurus genomic clone
CH240_298D10, genomic survey sequence.
ACCESSION
CC474046
VERSION
CC474046.1 GI:31751163
SOURCE
GSS.
Bos taurus (cow)
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



genes). The size range of inserts for the library was 0.5Kb - 2.5 Kb, the cDNA being directionally cloned between the EcoRI and the XhoI sites of the Bluescript phagemid. Bluescript phagemid was excised by mass in vivo excision of the library and inserted into E.coli XL1 Blue. The plasmid was purified using the Qiagen QiaWell 8 system and sequenced with Taq dyeterminator chemistry (SK,T3 and T7 primers) for ABI 373A. The library was prepared by Hanan Abdel Hamed, AIn Shams Univ., Cairo and provided by Mohamed Saber, Theodore Bilharz Research Institute, Cairo."

## ORIGIN

Query Match 2.0%; Score 23; DB 14; Length 259;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AATATATATATATATACAAATTA 657

DB 36 AATATATATATATATACAAATTA 58

## RESULT 10

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112  
EST93059 Skin tumor I Homo sapiens cDNA 5' end, mRNA sequence.  
AA380112  
AA380112.1 GI:2032451

Homo sapiens (human)

## REFERENCE

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

## AUTHORS

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA380112/c 307 bp mRNA linear EST 21-APR-1997  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

/db\_xref="taxon:9606"

/dev\_stage="adult"

/clone\_lib="Skin tumor I"

/note="Organ: skin; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 307;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1085 AAAAGATTGGAAAGATTTCAT 1107

DB 268 AAAAGATTGGAAAGATTTCAT 246

## RESULT 11

CB378526 449 bp mRNA linear EST 16-MAY-2003  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

CB378526

## FEATURES

## source

1. .449  
/organism="Heterodera glycines"  
/mol\_type="mRNA"  
/db\_xref="taxon:51029"  
/sex="mixed"  
/tissue\_type="whole organism"  
/dev\_stage="3rd stage juvenile"  
/lab\_host="DH10B"  
/clone\_lib="Heterodera glycines J3"  
/note="Vector: pBluescript SK+ (Stratagene); Site 1: XhoI;  
Site 2: EcoRI; This library was generated by cloning cDNAs  
directionally into Uni-ZAP(Stratagene) (T3 primer/EcoRI  
are at the 5'-end and T7/XhoI are at the 3'-end). The  
library was excised [now in pBluescript SK(+)] and  
normalized (Bonaldo et al 1996 Genome Research 6:791-806).  
Library constructed by Thomas Baum (tbaum@iastate.edu),  
Iowa State University, Plant Pathology Department and Jeff  
McDermott (jmcderm@iastate.edu).  
Putative full length read  
Seq primer: T3 from Gibco  
High quality sequence stop: 449.  
Location/Qualifiers



Iowa State University, Plant Pathology Department and Jeff McDermott (jpmcderm@iastate.edu)."

```

ORIGIN
Query Match      2.0%; Score 23; DB 14; Length 449;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 AACAAACAAATTTATGTTTCAT 82
|||||
Db 399 AACAAACAAATTTATGTTTCAT 421

RESULT 12
TA191E02P      543 bp      DNA      linear      GSS 13-DEC-2000
LOCUS          T. brucei sheared genomic DNA clone 191e02, forward sequence,
DEFINITION     genomic survey sequence.
ACCESSION      AL478092
VERSION        AL478092.1 GI:11841902
KEYWORDS       GSS.
SOURCE         Trypanosoma brucei
ORGANISM       Trypanosoma brucei
                Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
                Trypanosoma.
REFERENCE      1 (bases 1 to 543)
AUTHORS        Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
                Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
                Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE          Direct SubMISSION
JOURNAL        Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
                project, Sanger Centre, The Wellcome Trust Genome Campus,
                Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
                nh@sanger.ac.uk
COMMENT        Constructed at the Institute for Genomic Research (TIGR),
                Rockville, MD. Genomic DNA isolated from a cloned population of
                Trypanosoma brucei (TREU27/4 Gutat 10.1) was mechanically sheared
                to give a tight size distribution (
                4 kb). The v + i method used for the library construction is
                described in detail in Smith, R. and Venter, J.C. (Making small
                insert libraries for whole genome shotgun sequencing projects. In
                Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
                Barrell, Oxford University Press, 1999).
                Email: nh@sanger.ac.uk
                Details of T. brucei sequencing at the Sanger Centre are available
                at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES       source
                1. 543
                /organism="Trypanosoma brucei"
                /mol_type="genomic DNA"
                /strain="TREU27"
                /db_xref="taxon:5691"
                /clone="191e02"

ORIGIN
Query Match      2.0%; Score 23; DB 29; Length 543;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 535 AATATATATATATATACAAATTA 657
|||||
Db 201 AATATATATATATATACAAATTA 223

RESULT 13
CB607569      604 bp      DNA      linear      GSS 29-SEP-2003
LOCUS          tigr-gss-dog-17000312937745 Dog Library Canis familiaris genomic,
DEFINITION     genomic survey sequence.
ACCESSION      CB607569
VERSION        CB607569.1 GI:36924408
KEYWORDS       GSS.
SOURCE         Canis familiaris (dog)

ORIGIN
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 604)
Kirkness, S.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. 604
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match      2.0%; Score 23; DB 29; Length 604;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 490 AATACAAATATTTAAATATAT 512
|||||
Db 92 AATACAAATATTTAAATATAT 114

RESULT 14
CE223942      611 bp      DNA      linear      GSS 25-SEP-2003
LOCUS          tigr-gss-dog-17000326866999 Dog Library Canis familiaris genomic,
DEFINITION     genomic survey sequence.
ACCESSION      CE223942
VERSION        CE223942.1 GI:35379657
KEYWORDS       GSS.
SOURCE         Canis familiaris (dog)
ORGANISM       Canis familiaris
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
                1 (bases 1 to 611)
                Kirkness, S.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
                Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
                Venter, J.C.
                The dog genome: survey sequencing and comparative analysis
                Science 301 (5641), 1898-1903 (2003)
                22875432
                14512627
                Contact: Kirkness EF
                The Institute for Genomic Research
                Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
                Rockville, MD 20850, USA
                Tel: 301-838-0200
                Fax: 301-838-0208
                Email: ekirknes@tigr.org
                Class: shotgun.
                Location/Qualifiers
                1. 611
                /organism="Canis familiaris"
                /mol_type="genomic DNA"
                /strain="Standard Poodle"
                /db_xref="taxon:9615"

```

OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (750001)..(765000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (765001)..(780000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (780001)..(795000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (795001)..(810000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (810001)..(825000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (825001)..(840000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (840001)..(855000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (855001)..(870000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (870001)..(885000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (885001)..(900000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (900001)..(915000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc\_feature

Query Match 1.7%; Score 20; DB 4; Length 1230025;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 GAAGAAAAAGATTAATAAAA 122  
Db 1175674 GAAGAAAAAGATTAATAAAA 1175655

RESULT 12  
US-09-669-751-213  
; Sequence 213, Application US/09669751  
; Patent No. 6551575  
; GENERAL INFORMATION:  
; APPLICANT: Greenspan, Ralph J.  
; TITLE OF INVENTION: Methods for Identifying Compounds for  
; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to  
; TITLE OF INVENTION: Balance and the Perception of Gravity  
; FILE REFERENCES: P-NI 3864  
; CURRENT APPLICATION NUMBER: US/09/669,751  
; CURRENT FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US 60/168,579  
; PRIOR FILING DATE: 1999-12-02  
; NUMBER OF SEQ ID NOS: 261  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 213  
; LENGTH: 181  
; TYPE: DNA  
; ORGANISM: Drosophila  
US-09-669-751-213

Query Match 1.7%; Score 19; DB 4; Length 181;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 AAATATAACACACAAAA 396  
|||||

Db 107 AAATATAACACACAAAA 125

RESULT 13  
US-09-489-039A-2871/c  
; Sequence 2871, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2871  
; LENGTH: 1212  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2871

Query Match 1.7%; Score 19; DB 4; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 ACAATTAAATAAAAAAGGG 669  
|||||

Db 47 ACAATTAAATAAAAAAGGG 29

RESULT 14  
US-09-134-001C-391/c  
; Sequence 391, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 391  
; LENGTH: 1776  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-391

Query Match 1.7%; Score 19; DB 4; Length 1776;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 TTAATATAATTAATTGTAAT 192  
|||||

Db 547 TTAATATAATTAATTGTAAT 529

RESULT 15  
US-09-023-655-301  
; Sequence 301, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION

```

/ NUMBER OF SEQUENCES: 1508
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
/ STREET: 3174 PORTER DRIVE
/ CITY: PALO ALTO
/ STATE: CALIFORNIA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/023,655
/ FILING DATE: HEREWITH
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Zeller, Karen J.
/ REGISTRATION NUMBER: 37,071
/ REFERENCE/DOCKET NUMBER: PA-0001 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 301:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2590 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: LAINUT02
/ CLONE: 1353065
/ US-09-023-655-301

```

```

Query Match      1.7%; Score 19; DB 4; Length 2590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1068 AATAATAAATGAAGAAA 1086
    |||||||
Db 2265 AATAATAAATGAAGAAA 2283

```

```

Search completed: June 2, 2004, 07:16:20
Job time : 117 secs

```

1	1148	100.0	1148	6	ABN83916	ABn83916	Truncated
2	1148	100.0	1832	6	ABN83929	ABn83929	ABn83929
3	1122	97.7	1122	6	ABN83917	ABn83917	Truncated
4	1046	91.1	1821	6	ABD29066	ABn29066	Phaseolus
5	107	9.3	1866	6	ABN83928	ABn83928	Arceclin-4
6	107	9.3	3872	6	ABN83927	ABn83927	Arceclin-3
7	95	7.4	388	6	ABN83926	ABn83926	Arceclin-5
8	41	3.6	1211	2	AAQ94051	ABn94051	Kidney be
9	26	2.3	316	6	ABN83925	ABn63925	Arceclin-4
10	26	2.3	322	6	ABN83924	ABn83924	Arceclin-3
11	25	2.2	1542	1	AAJ90025	ABn90025	DNA encod
12	24	2.1	24	7	ABZ80122	ABz80122	Arceclin 5
13	24	2.1	1224	6	AA562538	AA562538	CDNA sequ
14	23	2.0	400	4	ABAA6475	ABa6475	Human bre
15	23	2.0	400	4	AAK30701	AAk30701	Human bon
16	23	2.0	400	4	AAK05134	AAk05134	Human bra
17	23	2.0	5067	9	ADP48876	ADd48876	Human gen
18	22	1.9	469	5	AA543419	AAa43419	Human G-P
19	22	1.9	2453	5	ABU21668	ABj21668	Drosophil
20	22	1.9	7374	6	ABK40036	ABk40036	Human che
21	22	1.9	7374	6	ABJ70602	ABj70602	Chemical
22	22	1.9	10080	3	AAJ95933	AAa95933	Human KIK
23	22	1.9	14312	4	AAK87515	AAk87515	Human imm



Db 719 AAAAAAATAATTTATGTTTCATTTGAGATTAGGAAGTAAAGAAATAAATAA 778  
Qy 121 AATATGTCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180  
Db 779 AAAATGTCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 838  
Qy 181 AATAATTTGTAATTTAGGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
Db 839 AATAATTTGTAATTTAGGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 898  
Qy 241 CAATTTTGTTHAACAATCTTATTAAGAACTTTAGTTAGTTAGTTAGTTAGTT 300  
Db 899 CAATTTTGTTHAACAATCTTATTAAGAACTTTAGTTAGTTAGTTAGTTAGTT 958  
Qy 301 AAAAAAATAACACAGAGAAACAAATTAACCCCTACTACCGTACCGTACCGTAC 360  
Db 959 AAAAAAATAACACAGAGAAACAAATTAACCCCTACTACCGTACCGTACCGTAC 1018  
Qy 361 AATGTTTGTATATCAATTAATAATTAACACACAAATAATCACTAATTAATTA 420  
Db 1019 AATGTTTGTATATCAATTAATAATTAACACACAAATAATCACTAATTAATTA 1078  
Qy 421 GTTATACATATTTTGTAAACCTTTAGAGTTTTCAAAACATTTCTAATACAT 480  
Db 1079 GTTATACATATTTTGTAAACCTTTAGAGTTTTCAAAACATTTCTAATACAT 1138  
Qy 481 AGTTTATAGAAATACAAATATTTAAATAATTAATTTTAAATAATTTCTAAAGT 540  
Db 1239 AGTTTATAGAAATACAAATATTTAAATAATTAATTTTAAATAATTTCTAAAGT 1198  
Qy 541 CAGATCTCTCACACCTCTGTGATCATTTTGTCTATGTATGTATGTATGTATGT 600  
Db 1199 CAGATCTCTCACACCTCTGTGATCATTTTGTCTATGTATGTATGTATGTATGT 1258  
Qy 601 ACAACAGAGTAAATAAATAAGGTAATACTAGGGAATATATATATATATATAT 660  
Db 1259 ACAACAGTAAATAAATAAGGTAATACTAGGGAATATATATATATATATATAT 1318  
Qy 661 AAAAAAGGAAATCAATTAAGATTTTGTATTTTGTATTTTGTATTTTGTATTT 720  
Db 1319 AAAAAAGGAAATCAATTAAGATTTTGTATTTTGTATTTTGTATTTTGTATTT 1378  
Qy 721 GCTGCCACTCAGCTCCCTCTCTCCACACATGCTCTCATGTCTCTCTCTCTCT 780  
Db 1379 GCTGCCACTCAGCTCCCTCTCTCCACACATGCTCTCATGTCTCTCTCTCTCT 1438  
Qy 781 TTCATATGACACAACTGCCATGATGTCGATGATGTCGATGATGTCGATGATG 840  
Db 1439 TTCATATGACACAACTGCCATGATGTCGATGATGTCGATGATGTCGATGATG 1498  
Qy 841 GACACCACTGGCATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 900  
Db 1499 GACACCACTGGCATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 1558  
Qy 901 TGGCCATGACACATGCCACCTCAGACCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
Db 1559 TGGCCATGACACATGCCACCTCAGACCTCTCTCTCTCTCTCTCTCTCTCTCT 1618  
Qy 961 CTCTCTCTCAATAATATCTATTTAAATTAACATTAATTTATATATATATATAT 1020  
Db 1619 CTCTCTCTCAATAATATCTATTTAAATTAACATTAATTTATATATATATATAT 1678  
Qy 1021 CGTGATGATCTGCCATGCTTTGTTTAAATAATTTAAATAATTTGAGTTGAGTT 1080  
Db 1679 CGTGATGATCTGCCATGCTTTGTTTAAATAATTTAAATAATTTGAGTTGAGTT 1738  
Qy 1081 AAGAAAAAGTTGGAAGATTTTGTATTTGTATTTGTATTTGTATTTGTATTTGT 1140  
Db 1739 AAGAAAAAGTTGGAAGATTTTGTATTTGTATTTGTATTTGTATTTGTATTTGT 1798  
Qy 1141 TTAATGCA 1148  
Db 1799 TTAATGCA 1806

RESULT 3  
ABN83917  
ID AEN83917 standard; DNA; 1122 BP.

XX AC AEN83917;

XX AC AC

XX DT 06-SEP-2002 (first entry)

XX DE Truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter #2.

XX KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;

XX OS pharmaceutical; ds.

XX OS Phaseolus vulgaris.

XX PN W0200250295-A2.

XX PD 27-JUN-2002.

XX PF 17-DEC-2001; 2001WO-US047495.

XX PR 18-DEC-2000; 2000US-0255979P.

XX PA (RENE-) RENESSEN LLC.

XX PI Wang Q, Dubois P, Liang J, Oulmassov T;

XX DR WPI; 2002-508809/54.

XX PT New transformed or transgenic soybeans plants or cells with an Arcelin-5 promoter, for use as an improved dietary source of protein for humans or animals, or for producing soybeans with important agricultural or nutritional properties.

XX PS Example 5; Page 66; 74pp; English.

XX CC The invention relates to a transformed soybean plant cell and transgenic soybean plant, both of which has a nucleic acid molecule comprising the Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence. The transformed soybean plant cell and transgenic soybean plant are useful as an improved source of dietary protein for humans and livestock. These are also useful for producing soybean plants that exhibit important agricultural, nutritional or pharmaceutical properties. The current sequence represents a truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter sequence. This sequence was used in the expression of a GUS reporter gene in a soybean cotyledon

XX SQ Sequence 1122 BP; 410 A; 210 C; 146 G; 356 T; 0 U; 0 Other;

Query Match

Best Local Similarity 97.7%; Score 1122; DB 6; Length 1122;

Matches 1122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCTCTCAATAGAAAATGTGTTATTTCTCTATCAGACAAAGGGGACAGTTAAC 62

Db 1 GATCTCTCAATAGAAAATGTGTTATTTCTCTATCAGACAAAGGGGACAGTTAAC 60

Qy 63 AATCTCTTATGTTTTCATTTCAGATTAGGAGGTAAGGTAAGGTAAGGTAAGGTA 122

Db 61 AATCTCTTATGTTTTCATTTCAGATTAGGAGGTAAGGTAAGGTAAGGTAAGGTA 120

Qy 123 AATCTCTTATGTTTTCATTTCAGATTAGGAGGTAAGGTAAGGTAAGGTAAGGTA 182

Db 121 AATCTCTTATGTTTTCATTTCAGATTAGGAGGTAAGGTAAGGTAAGGTAAGGTA 180

Qy 183 TAATCTTAATAGGTTTTCAGTTCAGGACCACTCAGAGCAAGATTTCAAGAAACA 242

Db 181 TAATCTTAATAGGTTTTCAGTTCAGGACCACTCAGAGCAAGATTTCAAGAAACA 240

Qy 243 AATCTCTTAACATCTTATTTAGAAACTTTTGTAGTTAGTTAGTTAGTTAGTTAG 302

Db 243 AATCTCTTAACATCTTATTTAGAAACTTTTGTAGTTAGTTAGTTAGTTAGTTAG

241	Db	ATTTTGTAAACATCTTATTAGAAACCTTTAGTTAGCTTGTGAAGTTAGTAATTAACAAA	300
303	Qy	AAAAATTACACACGAGAGAAACACAAATAAACCCACTACCGTCAGGTTATCATPAAGATGAAA	362
301	Db	AAAAATTACACACGAGAGAAACACAAATAAACCCACTACCGTCAGGTTATCATPAAGATGAAA	360
363	Qy	TGTTTTGATATCAATTAATAATAACACACACAAAAAAATACATCTAAATTATTAACAATATATGT	422
361	Db	TGTTTTGATATCAATTAATAATAACACACACAAAAAAATACATCTAAATTATTAACAATATATGT	420
423	Qy	TATACATATATTTTTGTAAAAACCTTAGAGTTTTTCAAAACATCTTAATACATCATTTAGAG	482
421	Db	TATACATATATTTTTGTAAAAACCTTAGAGTTTTTCAAAACATCTTAATACATCATTTAGAG	480
483	Qy	TTTATAGAANAACAAATATTTAAAAAATAATAATTTTAAAAAAAACATCTTAAGCTATTCA	542
481	Db	TTTATAGAANAACAAATATTTAAAAAATAATAATTTTAAAAAAAACATCTTAAGCTATTCA	540
543	Qy	GATCCTCTCACACTGTGTGATCACTTTAGTCATGTATGTAGTACATCATTTGTAGTTCAC	602
541	Db	GATCCTCTCACACTGTGTGATCACTTTAGTCATGTATGTAGTACATCATTTGTAGTTCAC	600
603	Qy	AACAGAGTAAAAATAATAAGCATAAACTAGGCAATATATATATATATACAATTAATAATA	662
601	Db	AACAGAGTAAAAATAATAAGCATAAACTAGGCAATATATATATATATACAATTAATAATA	660
663	Qy	AAAAAGGAAAAATCAAAATTAGATTTTTTGATTTCCCAATGACACAACTGACCATGCACGC	722
661	Db	AAAAAGGAAAAATCAAAATTAGATTTTTTGATTTCCCAATGACACAACTGACCATGCACGC	720
723	Qy	TGCCACCTCAGCTCCCTCCCTCTCCACACATGTCTCATGTCACTTTCGACTTTGGCTTTTT	782
721	Db	TGCCACCTCAGCTCCCTCCCTCTCCACACATGTCTCATGTCACTTTCGACTTTGGCTTTTT	780
783	Qy	CACATGACACAACTCCCATGCAATGTTGCCACGTGAGCTCCTTCCTCCCATGATGA	842
781	Db	CACATGACACAACTCCCATGCAATGTTGCCACGTGAGCTCCTTCCTCCCATGATGA	840
843	Qy	CACCACTGGGCACTGCATGCTGCCACCTCAGCTCCCACTCTTCTCATATGAGCCTACTG	902
841	Db	CACCACTGGGCACTGCATGCTGCCACCTCAGCTCCCACTCTTCTCATATGAGCCTACTG	900
903	Qy	GCCATGACACTGCCACTCAGCACTCCCTCTCACTTCCCACTGCTACCTGCCAAACCGCT	962
901	Db	GCCATGACACTGCCACTCAGCACTCCCTCTCACTTCCCACTGCTACCTGCCAAACCGCT	960
963	Qy	TCTCTCCATAAATATCTATTTAAATTTAAACTAATTTTCATATACCTTTTTTGATGACG	1022
961	Db	TCTCTCCATAAATATCTATTTAAATTTAAACTAATTTTCATATACCTTTTTTGATGACG	1020
1023	Qy	TGGATGCAATGGCCATCGCTGTTTAATAATTTGTTAAATTTGGAGTTGAATAATAAATGAAA	1082
1021	Db	TGGATGCAATGGCCATCGCTGTTTAATAATTTGTTAAATTTGGAGTTGAATAATAAATGAAA	1080
1083	Qy	GAIAAAAAAGTTGGAAAGATTTTGCAATTTGTTGTGTATAAATA	1124
1081	Db	GAIAAAAAAGTTGGAAAGATTTTGCAATTTGTTGTGTATAAATA	1122

RESULT 4

RESUL 4  
AAD29066

AAD29066  
ID AAD29066 standard; DNA: 1821 BP.

XX  
990627WY DT

AC AAD29066:

XX  
7906270000

XX 07-MAY-2002 (first entry)

XX  
ID 2007-TM-10 \ PSTT)

DE Phaseolus vulgaris arcelin promoter.

DE  
XX  
FIASCOIUS VUJGALLS arcetrii promoter

Heterologous gene expression: plant: arcelin promoter: arcelin leader:

RW heterologous gene expression; ds.  
 KW seed preferred expression cassette: ds.

XX  
NY  
naas natatard' paas  
xxxxxxxxxxxxxxxx

*Phaseolus vulgaris*.

US Phaeotus vulgaris.

[illegible]

WO200200899-A2.  
03-JAN-2002.  
31-MAY-2001; 2001WO-EP006298.  
29-JUN-2000; 2000EP-00202278.  
(VLA-) VLAMMS INTERUNIVERSITAIR INST BIOTECHNOG.  
Angenon G, De Jaeger G, Goossens A, Depicker A;  
WPI; 2002-139925/18.  
Novel seed preferred expression cassette having gene regulatory elements comprising arcelin promoter, arcelin 5l leader, and arcelin 5l 3' end, useful for heterologous gene expression in plants.  
Claim 1; Page 46-47; 52pp; English.  
The invention relates to heterologous gene expression in plants. The invention also relates to seed preferred expression cassette having gene regulatory elements comprising arcelin promoter, arcelin 5l leader, and arcelin 5l 3' end. This expression cassette is useful for heterologous gene expression in plants. The protein encoded by the heterologous gene is a single chain antibody variable fragment (scFv). The present sequence is Phaseolus vulgaris arcelin promoter of the invention  
Sequence 1821 BP; 698 A; 308 C; 210 G; 605 T; 0 U; 0 Other;  
Very Match 91.1%; Score 1046; DB 6; Length 1821;  
st Local Similarity 99.8%; Pred. No. 0;  
tches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
1 TAGGATCCTTCAATAGAAAATGTGTATTTCCTCATCCAGACAAAGGGCCACAGTTA 60  
674 TAGGATCCTTCAATAGAAAATGTGTATTTCCTCATCCAGACAAAGGGCCACAGTTA 733  
61 ACACAAACAAATTTATGTCTTCATTTGAGATTAAGGAAGGTAGGAAGAAAAAGATTAAAA 120  
734 ACACAAACAAATTTATGTCTTCATTTGAGATTAAGGAAGGTAGGAAGAAAAAGATTAAAA 793  
121 AAAATGTCCTTATCTCTTTGTTTCGTGAATATATATATAGAGACTTAACTTTTATAT 180  
794 AAAATGTCCTTATCTCTTTGTTTCGTGAATATATATATAGAGACTTAACTTTTATAT 853  
181 AATAATTTGTAATTTAGTGTTCCTAGTCATGAGCACCACTCAGAGACAAGATTTCAAGAAA 240  
854 AATAATTTGTAATTTAGTGTTCCTAGTCATGAGCACCACTCAGAGACAAGATTTCAAGAAA 913  
241 CAATTTTGTATAACATCTTATAGAAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAACA 300  
914 CAATTTTGTATAACATCTTATAGAAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAACA 973  
301 AAAAAAATATACACGAGAGAAACACAAATAAACCCACTACCGTCAGGTATCATAGAGTGA 360  
974 AAAAAAAGTACACACGAGAGACACAAATAAACCCACTACCGTCAGGTATCATAGAGTGA 1033  
361 AATGTTTGTATCATCAATTAATAATTAACACAAATAACATCTTAATTAACCAATATAT 420  
1034 AATGTTTGTATCATCAATTAATAATTAACACAAATAACATCTTAATTAACCAATATAT 1093  
421 GTTATACATATATTTTGTAAAAAACCCTAGAGTTTTTCAAAAACATTTCTAATACATGATTAG 480  
1094 GTTATACATATATTTTGTAAAAAACCCTAGAGTTTTTCAAAAACATTTCTAATACATGATTAG 1153  
481 AGTTTATAGAAATACAAATATTTTAAAAAATATATTTTAAAAAACAATCTCTAAGTCATT 540  
1154 AGTTTATAGAAATACAAATATTTTAAAAAATATATTTTAAAAAACAATCTCTAAGTCATT 1213  
541 CAGATCCTCTCACACCTGTGTGATCATTTAGTTCATGTATGTAGTACCAATCATTTGTAGTTC 600

```
Db 1214 CAGATCCTCTCACACCTGTGTGATCACTTAGTCATGTATGTAGTACAAATCATTGTAGTTC 1273
Qy 601 ACAACAGAGTAAATTAATAAAGGATAAACTAGGGAATATATATAATATAATATAATATAAT 660
Db 1274 ACAACAGAGTAAATTAATAAAGGATAAACTAGGGAATATATATAATATAATATAATATAAT 1333
Qy 661 AAAAAGGGAAATCAATTAAGATTTTGGATTTCCCAATGACACACAACTACCAATGCAC 720
Db 1334 AAAAAGGGAAATCAATTAAGATTTTGGATTTCCCAATGACACACAACTACCAATGCAC 1393
Qy 721 GCTGCCACCTCAGCTCCCTCTCTCCACACATGTCTCATGTGCACATTTGACATTTGGCTTT 780
Db 1394 GCTGCCACCTCAGCTCCCTCTCTCCACACATGTCTCATGTGCACATTTGACATTTGGCTTT 1453
Qy 781 TTCATATGACACAACTGCCATGCAATGTTCACAGTGCAGCTCCTCTCTCTCCATCAT 840
Db 1454 TTCATATGACACAACTGCCATGCAATGTTCACAGTGCAGCTCCTCTCTCTCCATCAT 1513
Qy 841 GACACACCTGGGCGATGCATGCTGCCACCTCAGCTCCCACTCTTCTCATTTATGAGCCTAC 900
Db 1514 GACACACCTGGGCGATGCATGCTGCCACCTCAGCTCCCACTCTTCTCATTTATGAGCCTAC 1573
Qy 901 TGGCAATGACACCTGCGACCTCAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db 1574 TGGCAATGACACCTGCGACCTCAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1633
Qy 961 CTCTCTCTCATAAATATCTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 1020
Db 1634 CTCTCTCTCATAAATATCTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 1693
Qy 1021 COTGATGATGATGCCATCGCTTCTTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 1080
Db 1694 COTGATGATGATGCCATCGCTTCTTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 1753
Qy 1081 AAGAAAGATTTGGAAGATTTTGCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
Db 1754 AAGAAAGATTTGGAAGATTTTGCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1813
Qy 1141 TTAATGCA 1148
Db 1814 TTAATGCA 1821

RESULT 5
ABN83928
ID ABN83928 standard; DNA; 1866 BP.
XX
AC ABN83928;
XX
DT 06-SEP-2002 (first entry)
XX
DE Arcelin-4 full length promoter sequence.
XX
KW Arcelin-4; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
PN WO200250295-A2.
XX
PD 27-JUN-2002.
XX
PF 17-DEC-2001; 2001WO-US047495.
XX
PR 18-DEC-2000; 2000US-0255879P.
XX
PA (RENE-) RENESSEN LLC.
XX
PI Wang Q, Dubois P, Liang J, Oulmassov T;
XX
PI WPI; 2002-508809/54.
XX
PS New transformed or transgenic soybeans plants or cells with an Arcelin-5
```

```
PT promoter, for use as an improved dietary source of protein for humans or
PT animals, or for producing soybeans with important agricultural or
XX nutritional properties.
XX Example 1; Fig 4; 74pp; English.
XX
CC The invention relates to a transformed soybean plant cell and transgenic
CC soybean plant, both of which has a nucleic acid molecule comprising the
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
CC The transformed soybean plant cell and transgenic soybean plant are
CC useful as an improved source of dietary protein for humans and livestock.
CC These are also useful for producing soybean plants that exhibit important
CC agricultural, nutritional or pharmaceutical properties. The current
CC sequence represents an arcelin-4 full length promoter sequence
XX
XX Sequence 1866 BP; 692 A; 339 C; 209 G; 626 T; 0 U; 0 Other;
XX
Query Match 9.3%; Score 107; DB 6; Length 1866;
Best Local Similarity 100.0%; Pred. No. 1.9e-38;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 862 TGGCACCTCAGCTCCCACTCTCTCTCATTTATGAGCCTACTGGCCATGCACACTGCCACT 921
Db 1544 TGGCACCTCAGCTCCCACTCTCTCTCATTTATGAGCCTACTGGCCATGCACACTGCCACT 1603
Qy 922 CAGCACTCCTCTCAGCTTCCCACTTGTCTACCTGCCAAACGGCTTCTCTC 968
Db 1604 CAGCACTCCTCTCAGCTTCCCACTTGTCTACCTGCCAAACGGCTTCTCTC 1650

RESULT 6
ABN83927
ID ABN83927 standard; DNA; 1872 BP.
XX
AC ABN83927;
XX
DT 06-SEP-2002 (first entry)
XX
DE Arcelin-3 full length promoter sequence.
XX
KW Arcelin-3; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
PN WO200250295-A2.
XX
PD 27-JUN-2002.
XX
PF 17-DEC-2001; 2001WO-US047495.
XX
PR 18-DEC-2000; 2000US-0255879P.
XX
PA (RENE-) RENESSEN LLC.
XX
PI Wang Q, Dubois P, Liang J, Oulmassov T;
XX
PI WPI; 2002-508809/54.
XX
PS New transformed or transgenic soybeans plants or cells with an Arcelin-5
PT promoter, for use as an improved dietary source of protein for humans or
PT animals, or for producing soybeans with important agricultural or
XX nutritional properties.
XX Example 1; Fig 4; 74pp; English.
XX
CC The invention relates to a transformed soybean plant cell and transgenic
CC soybean plant, both of which has a nucleic acid molecule comprising the
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
CC The transformed soybean plant cell and transgenic soybean plant are
CC useful as an improved source of dietary protein for humans and livestock.
CC These are also useful for producing soybean plants that exhibit important
CC agricultural, nutritional or pharmaceutical properties. The current
XX sequence represents an arcelin-4 full length promoter sequence
```



CC sequence represents an arcclin-3 full length promoter sequence  
XX  
SQ Sequence 1872 BP; 694 A; 339 C; 209 G; 630 T; 0 U; 0 Other;

Query Match 9.3%; Score 107; DB 6; Length 1872;  
Best Local Similarity 100.0%; Pred. No. 1.8e-38;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 862 TGGCAGCTCAGCTCCACCTCTTCTCATATGAGCCCTACTGGCCATGCACACTGCCACCT 921  
DB 1544 TGGCAGCTCAGCTCCACCTCTTCTCATATGAGCCCTACTGGCCATGCACACTGCCACCT 1603

QY 922 CAGCAGCTCTCTCAGCTCCACCTCTTCTCATATGAGCCCTACTGGCCATGCACACTGCCACCT 968  
DB 1604 CAGCAGCTCTCTCAGCTCCACCTCTTCTCATATGAGCCCTACTGGCCATGCACACTGCCACCT 1650

RESULT 7  
ABN83926  
ID AEN83926 standard; DNA; 288 BP.  
XX  
AC AEN83926;  
XX  
DT 06-SEP-2002 (first entry)  
XX  
DE Arcelin-5 promoter sequence fragment.  
XX  
KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;  
KW pharmaceutical; ds.  
XX  
OS Phaseolus vulgaris.  
XX  
PN WO200250295-A2.  
XX  
PD 27-JUN-2002.

XX  
PF 17-DEC-2001; 2001WO-US047495.  
XX  
PR 18-DEC-2000; 2000US-0255879P.  
XX  
PA (RENE-) RENESSEN LLC.  
XX  
PI Wang Q, Dubois P, Liang J, Oulmassov T;  
XX  
DR WPI; 2002-508809/54.  
XX  
PT New transformed or transgenic soybeans plants or cells with an Arcelin-5  
PT promoter, for use as an improved dietary source of protein for humans or  
PT animals, or for producing soybeans with important agricultural or  
PT nutritional properties.

XX  
PS Example 1; Fig 1; 74pp; English.  
XX  
CC The invention relates to a transformed soybean plant cell and transgenic  
CC soybean plant, both of which has a nucleic acid molecule comprising the  
CC phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.  
CC The transformed soybean plant cell and transgenic soybean plant are  
CC useful as an improved source of dietary protein for humans and livestock.  
CC These are also useful for producing soybean plants that exhibit important  
CC agricultural, nutritional or pharmaceutical properties. The current  
CC sequence represents an arcclin-5 promoter sequence fragment  
XX  
SQ Sequence 288 BP; 121 A; 26 C; 38 G; 103 T; 0 U; 0 Other;

Query Match 7.4%; Score 85; DB 6; Length 288;  
Best Local Similarity 100.0%; Pred. No. 2e-28;  
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1037 TCGTCTTTTAATAATTTGTTGAGTTGATTAATAATGAAAGAAAGTTGGAA 1096  
DB 204 TCGTCTTTTAATAATTTGTTGAGTTGATTAATAATGAAAGAAAGTTGGAA 263

QY 1097 AGATTTTGCATTTGTTGTGTATAA 1121

DB 264 AGATTTTGCATTTGTTGTGTATAA 288

RESULT 8  
AAQ94051  
ID AAQ94051 standard; DNA; 1211 BP.

XX  
AC AAQ94051;  
XX  
DT 10-MAY-1996 (first entry)  
XX  
DE Kidney bean lectin-like gene.  
XX  
KW lectin-like; kidney bean; insect resistance; genomic analysis;  
KW DNA marker; ss.

XX  
OS Phaseolus vulgaris.  
XX  
PN JP07132092-A.  
XX  
PD 23-MAY-1995.  
XX  
PF 11-NOV-1993; 93JP-00305988.  
XX  
PR 11-NOV-1993; 93JP-00305988.  
XX  
PA (NORQ) NORINSUISANSHO NOGYO SEIBUTSU SHIGEN.  
XX  
DR WPI; 1995-220115/29.

XX  
PT New kidney bean gene encoding lectin-like protein - for the production of  
PT insect-resistant crop plants.  
XX  
PS Claim 2; Fig 1-3; 6pp; Japanese.

XX  
CC The DNA encodes a lectin-like protein isolated from kidney bean var.  
CC Kentucky Wonder. The gene provides insect-resistance to crop plants, and  
CC is also useful as a DNA marker in gene and genomic analysis  
XX  
SQ Sequence 1211 BP; 344 A; 322 C; 203 G; 342 T; 0 U; 0 Other;

Query Match 3.6%; Score 41; DB 2; Length 1211;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1108 TTGTTGTTGTATAATAATAGAGAGAGAGAGATGCTGTTAATGCA 1148  
DB 263 TTGTTGTTGTATAATAATAGAGAGAGAGAGATGCTGTTAATGCA 303

RESULT 9  
ABN83925  
ID ABN83925 standard; DNA; 316 BP.  
XX  
AC ABN83925;  
XX  
DT 06-SEP-2002 (first entry)  
XX  
DE Arcelin-4 promoter sequence fragment.  
XX  
KW Arcelin-4; promoter; plant; transgenic; soybean; agriculture; nutrition;  
KW pharmaceutical; ds.

XX  
OS Phaseolus vulgaris.  
XX  
PN WO200250295-A2.  
XX  
PD 27-JUN-2002.  
XX  
PF 17-DEC-2001; 2001WO-US047495.  
XX  
PR 18-DEC-2000; 2000US-0255879P.

XX (RENE-) RENESSEN LLC.  
 XX Wang Q, Dubois P, Liang J, Oulmassov T;  
 PI WPI; 2002-508809/54.  
 DR  
 XX New transformed or transgenic soybeans plants or cells with an Arcelin-5  
 PT promoter, for use as an improved dietary source of protein for humans or  
 PT animals, or for producing soybeans with important agricultural or  
 PT nutritional properties.  
 PT  
 XX Example 1; Fig 1; 74pp; English.  
 XX  
 XX The invention relates to a transformed soybean plant cell and transgenic  
 CC soybean plant, both of which has a nucleic acid molecule comprising the  
 CC phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.  
 CC The transformed soybean plant cell and transgenic soybean plant are  
 CC useful as an improved source of dietary protein for humans and livestock.  
 CC These are also useful for producing soybean plants that exhibit important  
 CC agricultural, nutritional or pharmaceutical properties. The current  
 CC sequence represents an arcelin-4 promoter sequence fragment  
 XX  
 XX Sequence 316 BP; 113 A; 37 C; 37 G; 129 T; 0 U; 0 Other;  
 SQ  
 Query Match 2.3%; Score 26; DB 6; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 662 AAAAGGGGAAATCAATAGATT 687  
 DB 197 AAAAGGGGAAATCAATAGATT 222  
 RESULT 10  
 ABN83924  
 ID ABN83924 standard; DNA; 322 BP.  
 XX  
 AC ABN83924;  
 XX  
 XX 06-SEP-2002 (first entry)  
 DT  
 XX Arcelin-3 promoter sequence fragment.  
 DE  
 XX Arcelin-3; promoter; plant; transgenic; soybean; agriculture; nutrition;  
 KW pharmaceutical; ds.  
 XX  
 XX Phaseolus vulgaris.  
 OS  
 XX WO200250295-A2.  
 PN  
 XX 27-JUN-2002.  
 PD  
 XX 17-DEC-2001; 2001WO-US047495.  
 PF  
 XX 18-DEC-2000; 2000US-0255879P.  
 PR  
 XX (RENE-) RENESSEN LLC.  
 PA  
 XX Wang Q, Dubois P, Liang J, Oulmassov T;  
 PI WPI; 2002-508809/54.  
 DR  
 XX New transformed or transgenic soybeans plants or cells with an Arcelin-5  
 PT promoter, for use as an improved dietary source of protein for humans or  
 PT animals, or for producing soybeans with important agricultural or  
 PT nutritional properties.  
 PT  
 XX Example 1; Fig 1; 74pp; English.  
 PS  
 XX The invention relates to a transformed soybean plant cell and transgenic  
 CC soybean plant, both of which has a nucleic acid molecule comprising the  
 CC phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.  
 CC

CC The transformed soybean plant cell and transgenic soybean plant are  
 CC useful as an improved source of dietary protein for humans and livestock.  
 CC These are also useful for producing soybean plants that exhibit important  
 CC agricultural, nutritional or pharmaceutical properties. The current  
 CC sequence represents an arcelin-3 promoter sequence fragment  
 XX  
 XX Sequence 322 BP; 114 A; 37 C; 38 G; 133 T; 0 U; 0 Other;  
 SQ  
 Query Match 2.3%; Score 26; DB 6; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 662 AAAAGGGGAAATCAATAGATT 687  
 DB 197 AAAAGGGGAAATCAATAGATT 222  
 RESULT 11  
 AAN90025/c  
 ID AAN90025 standard; DNA; 1542 BP.  
 XX  
 AC AAN90025;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT  
 XX 17-DEC-2001 (revised)  
 DT  
 XX 01-NOV-1989 (first entry)  
 DE  
 XX DNA encoding antigenic circumsporozoite protein of Plasmodium malariae.  
 KW Antigenic; vaccine; Plasmodium malariae; circumsporozoite protein.  
 XX  
 XX Plasmodium malariae.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 CDS 98..1383  
 FT /\*tag= a  
 FT misc\_feature 424..1036  
 FT /\*tag= b  
 XX  
 XX USN7238746-N.  
 PN  
 XX 21-MAR-1989.  
 PD  
 XX  
 XX 21-MAR-1989; 89US-00238746.  
 PF  
 XX 31-AUG-1989; 88US-00238746.  
 PR  
 XX (USSH) NAT INST OF HEALTH.  
 PA  
 XX WPI; 1989-172958/23.  
 DR  
 XX  
 XX Antigenic protein for vaccination against malaria - encoded by cloned  
 PT gene coding for plasmodium malariae circumsporozoite protein.  
 PT  
 XX Disclosure; Fig 1; 11pp; English.  
 PS  
 XX  
 XX DNA encoding antigenic circumsporozoite protein of Plasmodium malariae.  
 CC Used as a vaccine against malaria. Nucleotides 424 - 1036 make up the  
 CC immunodominant region. (Note: Revised entry submitted to correct the  
 CC patent number format of US Government-owned NTIS applications to prevent  
 CC clashes with ongoing US granted patent numbers. For further information  
 CC please visit the Derwent web site at  
 CC www.derwent.com/dwpi/updates/ntis.us.html.) (Updated on 25-MAR-2003 to  
 CC correct PD field.) (Updated on 25-MAR-2003 to correct PP field.) (Updated  
 CC on 25-MAR-2003 to correct PR field.)  
 XX  
 XX Sequence 1542 BP; 618 A; 228 C; 344 G; 352 T; 0 U; 0 Other;  
 SQ  
 Query Match 2.2%; Score 25; DB 1; Length 1542;  
 Best Local Similarity 100.0%; Pred. No. 0.39;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 634 GAATATATATATATACAAATTAA 658

Db 1513 GAATATATATATATATACAAATTA 1489  
|||||  
RESULT 12  
ABZ80122  
ID ABZ80122 standard; DNA; 24 BP.  
XX AC  
XX ABZ80122;  
XX  
XX 22-MAY-2003 (first entry)  
XX  
XX Arcelin 5 promoter PCR primer SEQ ID NO:79.  
XX  
XX Gamma-tocopherol methyltransferase; methyltransferase; tocopherol;  
XX alpha-tocopherol; alpha-tocotrienol; stress; oxidative stress tolerance;  
XX oxygen; ozone; UV tolerance; ultra violet tolerance; cold tolerance;  
XX fungal pathogen; microbial pathogen; plant; enzyme; PCR primer; ss.  
XX  
XX Synthetic.  
XX  
XX WO2003016482-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 16-AUG-2002; 2002WO-US026047.  
XX  
XX 17-AUG-2001; 2001US-0312758P.  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
XX Van Eenennaam A, Valentin HE, Karunanandaa B, Hao M, Aasen E;  
XX Levering C;  
XX  
XX WPI; 2003-268314/26.  
XX  
XX Novel substantially purified tocopherol protein and nucleic acid encoding  
XX the same, useful for producing a plant having a seed with increased alpha  
XX -tocopherol level.  
XX  
XX Example 6; Page 94; 218pp; English.  
XX  
XX The present invention describes a substantially purified tocopherol  
XX protein (I). Also described is a method (M) for reducing expression of  
XX methyltransferase 1 (MT1) or gamma-tocopherol methyltransferase (GMT) in  
XX a plant, by transforming a plant with a nucleic acid molecule having an  
XX exogenous promoter region which functions in plant cells to cause the  
XX production of an mRNA molecule, and growing the transformed plant. (M) is  
XX useful for increasing the gamma-tocopherol content. A polynucleotide (II)  
XX encoding (i) can be used for producing a plant having a seed with an  
XX increased alpha-tocopherol or alpha-tocotrienol level. (II) is useful for  
XX modulating the reduction of the expression, expression, overexpression of  
XX (i) in a transformed plant to provide tolerance to a variety of stresses,  
XX e.g. oxidative stress tolerance such as to oxygen or ozone, UV tolerance,  
XX cold tolerance, or fungal/microbial pathogen tolerance. (II) is useful  
XX for obtaining other nucleic acid molecules or homologues from the same  
XX species, to screen cDNA or genomic libraries, to isolated promoters of  
XX cell enhanced, cell specific, tissue enhanced, tissue specific,  
XX developmentally or environmentally regulated expression profiles, as  
XX markers, for detecting single nucleotide polymorphisms, and to determine  
XX the level of (i) in a plant or pattern of expression of (i) encoded in  
XX part or whole by (II). The present sequence represents a PCR primer for  
XX an arcelin 5 promoter sequence, which is used in an example from the  
XX present invention  
XX  
XX Sequence 24 BP; 2 A; 12 C; 3 G; 7 T; 0 U; 0 Other;  
SQ  
Query Match 2.1%; Score 24; DB 7; Length 24;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 812 CCACGTGAGCTCTTCTCTTCCC 835  
|||||

Db 1 CCACGTGAGCTCTTCTCTTCCC 24  
RESULT 13  
AAS62538/c  
ID AAS62538 standard; CDNA; 1224 BP.  
XX AC  
XX AAS62538;  
XX  
XX 14-FEB-2002 (first entry)  
XX  
XX cDNA sequence #325 encoding novel human secreted protein.  
XX  
XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
XX immune deficiency disorder; blood disorder; inflammatory disorder;  
XX infectious disorder; gene therapy; antimicrobial; hepatotropic;  
XX immunosuppressive; antirheumatic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200177291-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 29-MAR-2001; 2001WO-US010485.  
XX  
XX 06-APR-2000; 2000US-0195604P.  
XX  
XX (GENY ) GENETICS INST INC.  
XX  
XX Wong GG, Clark HF, Fecthel K, Agostino MJ, Howes SH, Resnick RJ;  
XX Gulukota K, Graham JR;  
XX  
XX WPI; 2002-010900/01.  
XX  
XX New polynucleotides encoding secreted proteins useful for treating e.g.  
XX asthma, HIV and Crohn's disease.  
XX  
XX Claim 1; Page 245; 391pp; English.  
XX  
XX The present invention relates to the isolation of novel cDNA sequences  
XX which encode human secreted proteins. The cDNA sequences have been  
XX derived from a variety of human tissues. The invention also provides a  
XX method for producing proteins from these polynucleotide sequences. The  
XX proteins are useful for identifying compounds that modulate their  
XX activity and production, and the cell is also useful for identifying  
XX compounds that modulate expression of the polynucleotide sequences  
XX encoding the secreted proteins. The sequences of the invention are useful  
XX for treating diseases such as hyperproliferative disorders (e.g. cancer),  
XX immune deficiency disorders (e.g. severe combined immunodeficiency  
XX (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
XX (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and  
XX infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
XX the invention are also useful in gene therapy. AAS62214-AAS62838  
XX represent the cDNA sequences of the invention that encode for novel human  
XX secreted proteins  
XX  
XX Sequence 1224 BP; 395 A; 205 C; 204 G; 420 T; 0 U; 0 Other;  
SQ  
Query Match 2.1%; Score 24; DB 6; Length 1224;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 501 TTTAAAAAATATATTTTAAAAAA 524  
|||||  
Db 763 TTTAAAAAATATATTTTAAAAAA 740  
|||||  
RESULT 14  
ABA46475/c  
ID ABA46475 standard; DNA; 400 BP.  
XX AC  
XX ABA46475;  
XX

```
XX 01-FEB-2002 (first entry)
XX Human breast cell single exon nucleic acid probe #5170.
DE Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX Homo sapiens.
XX WO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000662.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX Claim 1; SEQ ID NO 5170; 327pp + Sequence Listing; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 400 BP; 89 A; 85 C; 94 G; 132 T; 0 U; 0 Other;
SQ
Query Match 2.0%; Score 23; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 490 AAATACAAATATTTAAAAAATAT 512
DB 250 AAATACAAATATTTAAAAAATAT 228
RESULT 15
AAK30701/c
ID AAK30701 standard; DNA; 400 BP.
XX
AC AAK30701;
XX 06-NOV-2001 (first entry)
```

```
XX Human bone marrow expressed single exon probe SEQ ID NO: 5258.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-489900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 5258; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX Sequence 400 BP; 89 A; 85 C; 94 G; 132 T; 0 U; 0 Other;
SQ
Query Match 2.0%; Score 23; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 490 AAATACAAATATTTAAAAAATAT 512
DB 250 AAATACAAATATTTAAAAAATAT 228
Search completed: June 2, 2004, 04:59:46
Job time : 547 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:15:29 ; Search time 4648 Seconds  
(without alignments)  
10705.207 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148  
Sequence: 1 tagatattctcaatagaaaa.....agagagtgtgtaatagca 1148

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rtd.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgt\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1148	100.0	1148	6	AX463281 Sequence
2	1148	100.0	1832	6	AX463294 Sequence
3	1144.8	99.7	1821	6	AX343909 Sequence
4	1144.8	99.7	3900	6	PVARG5X1
5	1122	97.7	1122	6	AX463282 Sequence
6	796.6	68.4	1866	6	AX463293 Sequence
7	795	69.3	1872	6	AX463292 Sequence
8	274.6	23.9	4564	8	PHVAKCIA
9	139.4	12.1	2288	8	AF193029
10	118.2	10.3	1768	8	PVPDLEC1
11	109.2	9.5	22243	3	PFVAR23A
12	108.4	9.4	1395	8	PVDLEC1
13	108.4	9.4	1441	8	PHVDLECA
14	95.8	8.3	24611	2	AC111404
15	95.2	8.3	93791	2	AC138073
16	95	8.3	125933	2	AC146394
17	92	8.0	1689	8	PHVLECT
18	92	8.0	4846	8	AF325188
19	91.8	8.0	111861	9	AC069435
20	91.8	8.0	170627	2	AC125567
21	91.6	8.0	1392	8	PVPDLEC2
22	91.2	7.9	175544	2	AC117342
23	90.6	7.9	258658	3	AE014832
24	90	7.8	8056	6	AX599046
25	90	7.8	110000	2	PFMAL8P1_12
26	89.8	7.8	810	8	AF592180
27	89.8	7.8	14867	3	AE001398
28	89.8	7.8	136688	9	AL353783
29	89.6	7.8	103344	9	HS1100E15
30	89.6	7.8	141275	2	EX510640
31	89.6	7.8	171333	5	EX248097
32	89.6	7.8	249867	2	AC127704
33	88.8	7.7	108908	3	PFMAL3P8
34	88.8	7.7	195620	2	EX088600
35	88.8	7.7	250029	3	AE014820
36	88.6	7.7	250743	3	AE014836
37	87.4	7.6	1434	8	AJ592058
38	87.4	7.6	2426	8	SDU49822
39	87.4	7.6	170880	2	EX640469
40	87.4	7.6	250029	3	AE014830
41	87.2	7.6	1453	8	AJ591978
42	87.2	7.6	164640	2	EX005461
43	87.2	7.6	168558	2	EX322549
44	87	7.6	3683	6	AX598999
45	87	7.6	136657	9	AC068600

## ALIGNMENTS

RESULT 1  
AX463281  
LOCUS AX463281 1148 bp DNA linear PAT 15-JUL-2002  
DEFINITION Sequence 1 from Patent WO0250295.  
ACCESSION AX463281  
VERSION AX463281.1 GI:21886232  
KEYWORDS Phaseolus vulgaris  
SOURCE Phaseolus vulgaris  
ORGANISM Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE 1  
AUTHORS Oulmassov, T., Wang, Q., Dubois, P. and Liang, J.

Wed

Db	794	AAA
Ov	181	ABT

CCCTTATCTCTTTGGTTTCTGTAAATAATAAAGAGACTTTAACTTTTAAATAT 853

794 181

RESULT 5	1122 bp	linear	PAT 15-JUL-2002
AX453282	DNA		
LOCUS			
DEFINITION	Sequence 2 from Patent WO0250295.		
ACCESSION	AX453282		

VERSION	AX463282.1	GI:21886233
KEYWORDS	Phaseolus vulgaris	
SOURCE	Phaseolus vulgaris	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.	
REFERENCE	1	
AUTHORS	Oulmasov, T., Wang, Q., Dubois, P. and Liang, J.	
TITLE	Arceclin-5 promoter and uses thereof	
JOURNAL	Patent: WO 0250295-A 2 27-JUN-2002;	
FEATURES	RENMESSEN LLC (US)	
source	Location/Qualifiers	
	1..1122	
	/organism="Phaseolus vulgaris"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:3885"	
ORIGIN		
	Query Match	97.7%; Score 1122; DB 6; Length 1122;
	Best Local Similarity	100.0%; Pred. No. 1.1e-166;
	Matches 1122; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	3	GGATCCTTCAATAGAAAATGTGTTATTTCTTCATCACAGACAAAGGGGCAACAGTTAAAC 62
Db	1	GGATCCTTCAATAGAAAATGTGTTATTTCTTCATCACAGACAAAGGGGCAACAGTTAAAC 60
QY	63	AAAAACAATTTATGTTTCATTCAGAGATTAAAGNAGGTTAAGGAGAGAAAAAGATTAAAAAA 122
Db	61	AAAAACAATTTATGTTTCATTCAGATTAAGGAGGTTAAGGAGAGAAAAAGAGATTAAAAAA 120
QY	123	AATGTCCTTATCTCTTTGTTTCTGTAATTAATATTAAGAGACTTAACCTTTTATATATA 182
Db	121	AATGTCCTTATCTCTTTGTTTCTGTAATTAATATTAAGAGACTTAACCTTTTATATATA 180
QY	183	TAATTTGTAATTAGGTTTCTTAGTCATGAGCACACATCTCAGAGACAAGATTTCAGAAAAACA 242
Db	181	TAATTTGTAATTAGGTTTCTTAGTCATGAGCACACATCTCAGAGACAAGATTTCAGAAAAACA 240
QY	243	ATTTTGTGTTAAACATCTTATTAGAAACCTTTTAGTTAGTCTCTGAAGTTAGAAATTAACAAA 302
Db	241	ATTTTGTGTTAAACATCTTATTAGAAACCTTTTAGTTAGTCTCTGAAGTTAGAAATTAACAAA 300
QY	303	AAAAATTCACACAGAGAAAACACAATAAACCCACTACCGTCAGGTTATCATAGAATGAAA 362
Db	301	AAAAATTCACACAGAGAAAACACAATAAACCCACTACCGTCAGGTTATCATAGAATGAAA 360
QY	363	TGTTTGTGATTCATTAAATATAACACACAAAAATACATCTAAATTAACAATATATGT 422
Db	361	TGTTTGTGATTCATTAAATATAACACACAAAAATACATCTAAATTAACAATATATGT 420
QY	423	TATACATATATTTTGTGAAAAACCTTAGAGTTTTTCAAAACATTCTAAATACATGATTAGAG 482
Db	421	TATACATATATTTTGTGAAAAACCTTAGAGTTTTTCAAAACATTCTAAATACATGATTAGAG 480
QY	483	TTTATAGAAATACAAATATTTTAAAAAATAATAATTTTAAAAAACATCTTAAAGTCATTCA 542
Db	481	TTTATAGAAATACAAATATTTTAAAAAATAATAATTTTAAAAAACATCTTAAAGTCATTCA 540
QY	543	GATCCTCTCACACCTGTGTCATCATTTAGTCATGTATGTAGTACAAATCATCTTGTAGTCCAC 602
Db	541	GATCCTCTCACACCTGTGTCATCATTTAGTCATGTATGTAGTACAAATCATCTTGTAGTCCAC 600
QY	603	TACAGAGTAAATAAATAAGGATTAACATAGGGATATATATAATATACATTAAATAA 662
Db	601	TACAGAGTAAATAAATAAGGATTAACATAGGGATATATATAATATACATTAAATAA 660
QY	663	AAAGGGAAAAATCAAAATTAGAAATTTTGTATGCCCACTGACACAACATCCCATCGACGC 722
Db	661	AAAGGGAAAAATCAAAATTAGAAATTTTGTATGCCCACTGACACAACATCCCATCGACGC 720
QY	723	TGCCACCTTCAGCTCCCTCCCTCTCCACACACATGTCTCATGTCACTTTTCGACTTGGCTTTT 782